The Vital Role of Metropolitan Access in Commuter, Regional, Intercity and Overnight Rail Passenger Transportation -- and Its Relationship to Technology

by Lexcie Lu
B.A., Natural Sciences (2000), University of Cambridge, England

Submitted to the Department of Civil Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Science in Transportation at the Massachusetts Institute of Technology

September 2003

Signature of Author ....
Center for Transportation Studies
August 31, 2003

Certified by ....
Carl D. Martland
Senior Research Associate
Thesis Supervisor

Accepted by ....
Oral Buyukozturk
Chairman, Committee for Graduate Students

[This page is intentionally left blank]

The Vital Role of Metropolitan Access in Commuter, Regional, Intercity and Overnight Rail Passenger Transportation -- and Its Relationship to Technology and Transit

by Lexcie Lu
B.A., Natural Sciences, University of Cambridge, England (2000).

Submitted to the Department of Civil Engineering in partial fulfillment of the requirements for the degree of Master of Science in Transportation at the Massachusetts Institvte of Technology

Abstract

The main thesis of this work is that planners and transportation professionals must think broadly in designing systems. Specifically, when designing an intercity transportation system, the objective is getting the customers from their actual origins to their ultimate destinations. With today's large and sprawling metropolitan areas, interconnections between urban and intercity transportation systems are a must – the customer's actual origin and ultimate destination are usually nowhere near an airport or a rail terminal. Whether this 'access leg' is provided by intermodal transfers or direct service is a matter of local circumstances, but it must be considered in the intercity transportation system planning process.

Future rail technology should not be designed to emulate either aircrafts or taxicabs. An aircraft is very good at traveling long distances quickly, but is unable to make intermediate stops, and thus a poor alternative for servicing dispersed demands. An automobile can make many intermediate stops efficiently, but cannot travel very fast. The ubiquitous automobile also suffers from ubiquitous urban congestion. Thus, it cannot service either extremely high demand densities or long corridors. Rail technology offers an intermediate option. In urban areas, rail offers efficient service to massive demands through high carrying capacity and dedicated rights-of-way. In rural areas it offers higher speeds by virtue of steel-wheels-on-steel-rails guidance. The combination of these two qualities makes intercity rail a winner in connecting one sprawling metropolitan area with another nearby – especially when coupled with such incremental enhancements as 'maglevication' of existing railroads. 'Shiny-go-faster' or personal rapid transit approaches ignore these advantages of rail transportation at their peril.

Intercity rail must exploit both advantages to compete effectively. The traditional, limited-stop high-speed rail approach ignores rail's ability to service many dispersed points of origin (streetcar suburbs), while the 'airport access' approach ignores the possibility of a direct service from a neighborhood 'subway station' to another one in a different metropolitan area. The key to success is not the one-seat-ride, but in eliminating the transfer, terminal and 'backtracking' time associated with many air-rail or air-bus solutions. These advantages are best demonstrated with a passenger utility model that is sensitive to the different values-of-time a customer perceives during difference phases of a door-to-door trip.

In the United States, higher speed rail is necessary in many cities for rail to stay competitive, but highest speeds are neither cost-effective nor necessary. Each scheme for increasing line-haul speed should be judged, using the total logistics-utility framework, against alternatives to improve access and options to make time disappear. Demands for speed, accessibility, amenities, and other upgrades that improve the customer utility must be balanced against each other. The results from the customer utility studies should be used to inform intercity transportation system design, to create a system that works in harmony to move people.

Thesis Supervisor:	Carl D. Martland
Title:	Senior Research Associate

Acknowledgements

Carl Martland is not only a thesis advisor, but also a good friend with the mandatory sense of humour and honesty, an effective internship broker, an encouraging instructor, a great teammate and debate partner, an effective coach. Thank you.

Nigel Wilson is no ordinary professor. Despite his heavy workload, he made time for a rogue British student whose thesis he wasn't supervising and had no responsibility for. Joe Sussman gave me sound advice, tolerated my sometimes brash manners with much patience, and funded me for the full two years with a project from Union Internationale de Chemins d'Fer. Some of other professors at MIT (and one at Harvard) provided live entertainment in the form of classes and saw me through the darkest moments: Cynthia Barnhart, Fredrick Salvucci, John Attanucci, Peter Belobaba, Henry Marcus, Ismail Chabini, George Kocur, Jose Gomez-Ibanez. I'm grateful to you all.

John Allen (MIT '91) got me into this big mess to start with, but still checked up on my progress, got me into writing for TRB, visited regularly, and helped with anything he could, including apartment hunting.

My friends were there to make sure I get through MIT. Angela, Erik, Shelli, Ryan, Heidi, Austin, Lissie, Alex, Kat, Ajay, Helen, Demian, Michelle, Joe, Tara, Jeff, Emily, Mike, Jodie, Donovan, Susie, Terence, and many others were great company, whether here or at a distance; for an intense moment or throughout. In particular, Erik's computer genius enabled things other than transportation to be discussed. Thank you for being friends.

Many people in the industry provided me with insight, data, interesting discussions, and opportunities to visit various facilities. Amongst them, Andreas Aeppli, Rob van Nes, Rohit Aggarwala, and Adam Rahbee were amongst the most helpful. I am grateful to each. Thanks are also due to the people of MBTA, the students in 1.011, and others at TRB who provided endless encouragement to reassure me that research doesn't always end up on dusty shelves.

The Department Secretaries make the place run and the place fun: Janet for cookie hour, Deena for alligator jokes, Pat for being my honourary Irish aunt, Ginny for medieval singsong, and others for everything else. Thank you for making MIT an experience, and a place I like. More broadly, I owe the State of Massachusetts and the United States much thanks, especially in the post September 11 world, for giving me an opportunity to develop and grow here.

My parents, Hongyang and Angela Lu, brought me into this world, got me an Intercity 125 trainset, gave me a first-class education, helped me get setup at MIT, made a point of talking to me to keep me sane, and kept supporting me even though I didn't become a rocket scientist. I couldn't have done it without you. Thank you for believing in me. As I started my thesis, Rachel, the love of my life, bore the brunt of my up-and-downs and supported me in more ways than imaginable, including by calling me a goober and making sure I ate properly. I couldn't have asked for a better partner in life. Thank you for putting up with me and making me happy.

Biographical Note

Photo by Sara Goplin, MIT Center for Transportation Studies

Grandson of a Formosan trucker, Lexcie Lu was born in the English textile city of Leeds -- a city served by both the London & North Eastern Railway and the London, Midland and Scottish. It is thus unsurprising that he developed a lifelong love for the railroads. As a foetus, he began his intercity commute between Bradford and Leeds on a British Rail Heritage diesel multiple-unit, since mom and dad hadn't found an apartment together yet.

True to the railroad tradition, he is a modern boomer who attended National Hsinchu Science & Technology Park Elementary School (1985-89), The Edinburgh Academy (1992-95), Felsted School (1995-97) before Trinity Hall, University of Cambridge (1997-2000) where he majored in Natural Sciences with specialization in Physics and Psychology. Before attending Massachusetts Institute of Technology, he was employed by two former British Rail companies: Railtrack (1998-99) and ScotRail (2000-01) in Glasgow, Scotland. Before he even finished his thesis, he decided that two years without setting foot on railroad property was too much, so he went to work for Massachusetts Bay Transportation Authority in Boston, even if it was only for 37 days.

Asked by his British friends why he moved, he jokes about moving from the Land of Pomp and Circumstance to the Land of Freedom and Truth. In reality, the first time he visited the United States, he was so impressed by the big trains and friendly people on Amtrak (Lake Shore Limited, 1998) that he decided he couldn't ride a British Rail Sleeper again and still think it's a real train. So he moved to be with the real trains.

[This page is intentionally left blank]

Contents

[This page is intentionally left blank]

List of Figures

List of Tables

[This page is intentionally left blank]

Chapter 1

Introduction

This thesis is a very general and very broad attempt at reframing the high-speed rail debate in the United States. This thesis represents a realization that intercity rail is a set of cross-cutting issues. Technological evaluation inevitably involves assumptions about network design and service planning, which in turn has implications on travel demand and institutional relationships between carriers operating different modes. In the longer term, infrastructure investment decisions and project evaluation would be influenced by all of the above. Focusing in any one area while assuming that all other areas would remain unchanged or continue the present trend, would inevitably result in a wrong answer. Worse still, the accumulation of small incremental changes (wrong answers) could result in the system evolving in a way that results in a state of affairs many would agree is undesirable.

The present research in many aspects of high-speed rail and other transportation options are too narrow in focus, and there is a strategic void in the role of rail in the future. At a policy level, there are many issues requiring discussion. Recently, issues such as the role of Amtrak, incremental high speed rail, intermodal connections, and transportation security have taken the center-stage in discussions of the future intercity ground transportation system. These are important issues, but these do not address the systems question – how would the U.S. deliver a transportation system that provides a reasonable service to a large number of people, at an affordable cost? The transportation system as a whole -- highways, railroads, airlines, urban transit, and other movement technologies -- should be thought of as different ways of getting people from one location to another. The transportation system user should be at the very center of this discussion: how would people want to travel from here to there? What is needed is not a supply-oriented approach that discusses modes, investment, institutions and service delivery, but a user-oriented approach that discusses market segments, their different needs, and the most economically efficient way to consolidate these needs so that services could be provided in bulk at lower costs. Warning: this approach would almost certainly cause disruptions in existing institutions, funding mechanisms, and modal coalitions.

At the heart of this thesis, there are a number of core ideas. When considering intercity passenger transportation, it is necessary to evaluate the utility of the entire trip experience from door to door -- similar to a total logistics cost model in freight transportation. When constructing a model of generalized costs, it is important to appreciate that different activities that occur during the trip have different values of time, depending both on the market segment and the quality of the experience. When evaluating high-speed rail projects or advanced guideway technologies, it is necessary to use this door-to-door utility model, applied from the customer's standpoint for a representative sample of origins and destinations, to ensure its success.

The United States is a country with a vastly distributed economy. Within the cities, suburbanization is rife -- people commute for up to 100 miles daily, and even within city neighborhoods the densities are low compared to many other countries. Between the cities, interstate commerce dominates most production operations as cities assume equally important roles in the region and the nation's economy. This type of economic geography calls for a totally-connected transportation network that takes people from one point directly to another, due to the general lack of 'supernodes' or natural hub locations. The result of applying this model suggests that provision of fast access from all locations is a necessary prerequisite to high-speed rail's success. Similarly, with important clusters of metropolitan areas being separated by large distances of rural plains, overnight rail could become competitive if service design were driven by a home-based generalized cost model, and amenities were made available but charged at marginal costs. In essence, overnight trains are able to 'pick up' within multiple cities in a region, travel overnight over areas of low demand density, and deliver people to multiple cities within the next industrial region.

Accumulation of small changes may result in important consequences. A historical example would be William Mulholland's great aquaducts, which brought more and more water into Southern California, but resulted in significant environmental damage elsewhere. The more sustainable solution would have been to include water demand management. Another example would be the phenomenon of urban sprawl -- which is wonderful in moderation, but dreadful when taken to extremes. The quest for ever higher speeds in rail transportation, is likely to be a similar wild goose chase, if maximum or average line-haul speed remain the only criteria for evaluation of high-speed rail schemes.

1.1 Thesis Outline

The total logistics cost model is necessary to understand the customer's journey experience. A total logistics perspective calculates utility as a function of traveler, trip, and transport characteristics, taking into account such things as the value of time for particular travelers and activities. The passengers are your customers from the moment they leave their home or work, and they remain your customers until they either arrive at the vacation resort, the home of the friend or family they are visiting, or back at their own home. Supporting the business traveler with hotels and local transportation at the destination city and providing them with information they need to navigate an unfamiliar city are all legitimate business objectives. Intercity transportation is a happy business; it's about taking care of people -- until they are safely in the hands of another trusted party, or no longer requires looking after. Thus, getting people from a transportation node to another node by the fastest possible means isn't necessarily in the best interest of an intercity carrier. This concept is reviewed in Chapter 2.

The value of time concerns the customer's experience during a particular point in the journey. Clearly, customers feel different about different experiences, and some may be willing to pay a premium to travel in a more pleasant environment. More importantly, amenities that the customers are willing to purchase may actually alter the perceived value-of-time. Although the morning commuter pays marginal costs to purchase a cup of coffee, the morning commute would be unbearable without it. If carriers were able to price onboard and en-route amenities at marginal costs, it can induce customers to behave in a way that lowers the perceived disutility of time spent en-route, indirectly influencing mode share. This idea is presented in Chapter 3.

Recent academic literature and state-of-practice on high-speed rail planning have mostly neglected the multi-faceted service dimension of the intercity transportation business. Technologies that are hailed as the next-generation high speed rail have mostly focused on increasing speed and reducing line haul time, sometimes at the expense of access and amenities. This approach is inappropriate, except in cases where the existing speeds are dismally slow (e.g. in a few very old parts of coastal and suburban corridors in the United States). Reducing line-haul time by skipping stops will increase total logistics costs for some passengers and may increase average logistics costs for all passengers if access time is properly accounted for. Increasing speed by reducing vehicle weight, reducing consist length, and introducing other such technical 'improvements' may actually increase the generalized logistics costs and diminish the value of the journey experience for the customer.

Some of the current high-speed rail schemes under consideration by state and Federal authorities could actually be a 'double whammy' when examined in light of this framework: sacrificing amenities for maximum speed not only increase the average passenger logistics cost for the market, but also fail to provide sufficient incremental benefits of journey time saving to offset the capital costs required for a new right-of-way or technology. With minimal amenities, the automobile with its low incremental cost per passenger, low out-of-pocket costs, easy access to amenities offered en-route by independent vendors, begins to look like an extremely attractive proposition. The 'shiny-go-faster' approaches that focus on maximum speed are at best not cost effective when examined in light of rational project evaluation and in some circumstances are a retrograde step for the average customer. Current plans and state of practice are analyzed in Chapter 4.

An important result when applying this framework to a region with sprawling metropolitan areas such as those in the United States is that the access time is actually a significant portion of the intercity travel time for the majority of customers. Thus, further reducing line-haul time may not achieve the desired goal. The most highly leveraged technologies in intercity rail travel may not be technologies that enable higher speeds on the mainlines, but those that allow many more stops with less time penalty per stop. Air technology is extremely apt at minimizing point-to-point travel time, but suffers from very poor access and inability to make multiple stops efficiently. Highway technology is limited to traveling at much lower speeds and suffers from the effect of urban congestion, but it is ubiquitous. Rail technology may find a niche by offering an intermediate level service that connects neighborhoods and suburbs in one metropolitan area directly to neighborhoods and suburbs in another metropolitan area. This idea is expostulated in Chapter 5.

Photo: Lexcie Lu & Mike Brotzman

Plate 1.1: Today's Sprawling Metropolitan Areas Require More Than One Intercity Transportation Access Point

The biggest payoff in the creation of this type of infrastructure is likely to be the enhanced commuter and regional rail access it will achieve, but intercity rail must take regional access seriously to withstand attack from both sides by air and auto. Contrary to what some public transit activists may believe, inability to offer one-seat rides is actually seriously hampering the competitiveness of regional and intercity rail in large metropolitan areas.

High speed rail advocates in the United States like to position overnight rail as the enemy. Those who are focused on increasing efficiency in a particular corridor portray overnight rail as outdated, outmoded, and an obstacle to efficient daytime train operations. Those who are focused on maximizing profitability realize that overnight rail has never been particularly profitable even in its heyday and prefer to focus management attention on shorter corridors that are both easier to manage and more profitable. These views are reasonable, but inconsistent with sound economic analysis. Although the demand for long-distance rail service is small, it is not zero, and long-distance trains are also able to carry local passengers unlike long-haul flights. Overnight rail fulfils an important niche in rail operations, and expands the portfolio of the rail carrier's offerings. Although it would never be the most profitable service, in partnership with airlines it could provide the day-return business traveler with a much upgraded service, while providing a late train for intermediate origins and destinations, and serves many other disparate purposes including express freight. The overnighter is a 'catch all' train. If its costs could be kept under control, continued operations of overnight trains will maintain network benefits for rail carriers and attract people living in rural areas to use rail network when they visit the metropolitan areas. A detailed evaluation of overnight services in the United States appears in Chapter 6.

1.2 Important Note

High speed rail investment proposals and evaluation criteria in the United States require a thorough rethink from the perspectives of potential customers. The current plans have mostly resulted from discussions between the operators, the informed current customers, and the politicians. Occasionally, non-customers enter into the discussion if the promoters threaten to take their house. Missing from the list of stakeholders are the millions in the United States who have never ridden a train and may not realize that trains still carried passengers. That's a big market to leave behind.

Using a theoretical model that evaluates passenger utilities for the entire door-to-door trip, this thesis demonstrates that the potential customers do not take the train for a good reason: the train does not meet their needs and is not competitive with the private automobile. Many live in a rural areas, but others live in suburbs around large metropolitan areas with rail service. Many of these are potential rail customers. For example, in the states traversed by the 'Capitol Flyer' – a hypothetical overnight service corridor running from Boston via Washington to Cleveland, Chicago and Milwaukee, approximately 40% of total population would live within a county with (or within 50 miles of) rail service. However, even within the Northeast Corridor in such origin-destination pairs as Washington-New York, both rail and air achieves a poor market share against the private auto. Longer door-to-door times compared to driving likely a handicap for collective intercity transportation in these areas.

Investing in access, amenities, and other improvements that increase the customer's total utility over the entire trip, from their departures from home until their arrivals back home, is likely to be as important – and much more cost-effective – than simply investing in speed. It is the customers – passengers – patrons – guests that we take care of, that ultimately make or break the railroad. We feed the customers, we keep them warm in the winter and cool in the summer, we give them material to read or provide other forms of entertainment, we give them directions and help with baggage. We price these services at marginal costs and we might then have some pricing power in terms of the fares. For instance, a New Jersey Transit study demonstrated that regular passengers in a commuter market would be willing to pay an additional $0.15 per trip for the convenience of having a luggage rack on board the vehicle. These amenities are much cheaper to provide than line-haul journey time reductions, but may have much more dramatic effect on revenue potential.

Intercity passenger transportation is a set of cross-cutting issues. Planners, carriers, and technology developers should be aware of these issues. Evaluation of intercity passenger rail proposals should take the wider, all-encompassing approach that may better represent the needs of potential customers who are not currently choosing rail.

Chapter 2

Transportation Demand and Utility Analysis Fundamentals

This chapter reviews the current state-of-practice in transportation demand modelling for both freight and passenger flows, and ridership forecasting. Implicitly, some generalized cost models (utility models, logistics cost models) are also reviewed as "feeder" to demand models. Predictive approaches that allow demand prediction sensitive to changes in transportation characteristics are favoured over explanatory approaches that simply correlate demand with non-transportation factors such as land-use and regional culture. The review of modelling methodologies is broad and all-emcompassing, but the comments on the current approaches will focus on adapting the methodologies for the purposes of high-speed intercity rail (HSR) planning and evaluation.

The chapter both demonstrate the wide variety of excellent approaches that have been developed in many different fields and highlights how application of the existing methodologies could be improved when specifically applied to medium-speed passenger transportation (i.e. 90mph~150mph). In general, the existing approaches encounter the following difficulties when applied to HSR: (1) inappropriate geographic aggregation, resulting in insufficient attention to access time; (2) insufficient detail with respect to utilities derived by passengers during various parts of the trip, resulting in a bias towards short and uncomfortable journeys; (3) failure to account for induced demand, resulting in a focus of investment in areas that are traditionally successful at the expense of rapidly developing areas.

The review of the many different demand modelling approaches may seem totally arbitrary and unrelated at first, however, each approach features strengths that could augment the development of an evaluation framework for HSR. Traditionally, urban transportation planning techniques have been applied to intercity transportation planning. However, adaptation of such models to HSR without due attention to the special conditions of the intercity travel market could result in misleading conclusions. Specifically, in the area of service planning, demand modelling issues must be well-understood to develop the optimal service plan.

2.1 Total Logistics Cost Models

Economists have studied the movement of goods for almost as long as goods have been moving. Not always in a formal way, but it is clear from the actions of the robber baron railroad tycoons in the late 19th century that they were aware of the changing value of goods due to spatial movement, and the inter-dependence between the cost-basis of the production facility and the economic well-being of the serving carrier (Schafer & Solomon, 1999). As evidenced by the rate-making structure that existed following the establishment of the Interstate Commerce Commission in 1887 for the regulation of railroads and other common carriers, it is clear that the notion of logistics costs were well understood by those in the industry. Higher-value items were subject to higher carrier tariffs, partially due to the firm's increased willingness-to-pay for the then-fastest mode of transportation, but also to account for the logistics costs that were avoided over slower modes such as canals and initially, the much higher costs of air transportation. This set-up for distributing the high capital costs of the railroad over many industries for essentially a shared facility functioned adequately for many years.

The basic ideas behind relating freight transportation demand to transportation costs, economic geography, and land-use patterns are very simple: the price charged by the retailer at the consumer's premises for given goods is made up of two components, the cost of production and the cost of transportation. This is called the spatial price equilibrium of commodities. Theoretical treatments of this concept at first considered the cost of production as fixed for each commodity, and the cost of transportation as linear with distance (Isard, 1956). More elaborate version of this model appeared subsequently, which suggested that transportation and production costs can be further fragmented into different components (Kresge & Roberts, 1970).

Production costs contain a component of land-rent, which changes as the location of production facilities are changed but does not change with the quantity produced; it contains a component of capital costs for production facilities, which does not vary with location in the long term (but will in the short term, when some costs are considered sunk); it also contains other terms which may or may not vary spatially and may or may not change with the quantity produced, such as the cost of labor and other inputs. Transportation costs can also be fragmented in the same way: there are fixed costs and variable costs, which may be affected by variables such as the transportation technology, transportation demand in a given corridor, and other variables. Other costs are also incurred in transit, such as warehousing costs, insurance costs against loss or damage to consignments, and time-value of money of capital tied up in goods-in-transit (called inventory costs). The basic thesis, thus, is that freight moves in such a way as to minimize total logistics costs. The total logistics costs is simply defined as the sum of all the aforementioned costs, summarized in the following equation (Roberts, 1971, 75).

Total Logistics Cost = Production cost + Ordering cost + Line-haul cost + Warehousing cost + Local delivery cost + Inventory cost + Risk + Stockout cost

Line-haul and local delivery costs are simply transportation costs associated with different modes (e.g. Line-haul cost might come from a rail freight bill, whilst Local delivery cost might be the cost of drivers and delivery vans); the warehousing cost is associated with the warehouse capacity requirements, and can possible be demand responsive (i.e. when goods compete for warehouse space, it is more expensive); the inventory cost is like the time-value of money (or value-of-time in passenger transportation), it is the cost associated with holding onto inventory and essentially paying interest on the value of the goods; the risk term is a way of quantifying the insurance cost against shipping and warehousing damage, which may be related to the mode or transport, the length of time that the inventory is held, and other factors such as the location of the warehouse. In an optimization model, the decision variables might be: the line-haul mode or carrier, the local delivery mode or carrier, the warehouse location and capacity, the minimum delivery quantity, and the inventory cost under different economic scenarios.

Given the abstract framework, Roberts formulated a utility model which predicts consumer choice based on extensive information about the available options (Roberts, 1975). In essence, the consumer utility is assumed to be equal to the total logistics costs, and relevant variables are changed until equilibrium is reached. If the objective is to evaluate choice of manufacturing sites, the quantity produced at each site is changed; if the objective is to evaluate whether to invest in a new technology for a specific corridor, the transportation cost or delay is changed. Sometimes, a joint-evaluation is required, since choices of production sites may depend on investment in new transportation facilities.

In a series of follow-up work, the total logistics cost idea (i.e. the utility model describing consumer preferences) was integrated with a number of different approaches to predicting transportation demand. A disaggregate choice model, combined with a logistics cost model, was used to develop a policy-sensitive working model for forecasting freight demand (Chiang, Roberts & Ben-Akiva, 1981). Empirical estimation of model parameters was carried out by building a database of intercity shipment flows, level-of-service attributes, commodity attributes, receiver attributes, and market attributes. The model allowed the practitioners to test the sensitivity of transportation demand with respect to a vast array of parameters and derive useful insight about the transportation system, however, the report acknowledges that there are innumerable ways in which the statistics can be improved though the better use of data. Carrier system concepts, combined with a total logistics cost model, gave rise to a decision support model that enabled carriers to leverage more value from high speed service, and shippers to make more optimal decisions amongst bids from different carriers and modes (Sheffi, 1988). Here, an optimization model is used to find the lowest cost for getting a piece of freight from the factory to the shopfloor.

Interestingly, the freight literanture is much closer to total logistics analysis than the passenger literature, possibly due to the traditionally vertically-integrated institutional arrangements in the freight industry (i.e. the same freight consolidator would take your freight at the origin and ensure it gets to the destination with one price). Models in the urban and air passenger transportation have traditionally had utility functions which related to the passenger's total logistics costs in much less detail than the freight models outlined above. Application of the total logistics concept to passenger flows will produce some results which have been documented only in a handful of cases, and somewhat contrary to conventional models (although entirely consistent with industry experience). Some models in the urban and air transportation sectors will be reviewed later in the chapter.

2.2 Classic Passenger Transportation Demand Models

In Manheim's classic Transportation Systems Analysis text (1979), a top-down approach is taken to understand transportation demand. Manheim describes transportation demand as consequences of long-run choices such as locational choices, activity patterns, and lifestyle aspirations. Having acknowledged that it may not be entirely possible to separate the long-run decisions from the short-run choices, he presents a utility model to predict the short-run portion of consumer behaviour. The basic assumption behind the model is that the long-run choices remains fixed. Given that, how would an individual decide which mode to take for a given trip on a given day?

Manheim suggests calculating passenger utility for a given mode (or combination of modes) by assuming that the utility is a function of a number of modal characteristics such as in-vehicle time, access time, out-of-pocket costs, service frequency, and others. Given the utility of a particular mode or combination of modes (an intermodal path), the utilities are then fed into a discrete choice model which computes the probability that one individual would choose a specific one of the given options for a given trip. Thus, the discrete choice model acts like a decision rule: given the relative characteristics of the modes or paths, how likely am I to choose mode *n* over all others? Many different formulations of the discrete choice model are possible. The logit model assumes that the probabilities are distributed with respect to utility according to the logistic function. The probit model assumes that the utilities for each path is randomly distributed about the mean value calculated, and asks what the probability of the utility of mode *n* (drawn from the normal distribution) is the highest of all modes. The assumption is then made that the total demand multiplied by the probability of a random individual choosing mode *n* would predict the total ridership expected on that mode (or that path).

There are three classic criticisms of this model. First, it assumes that transportation demand is fixed, and changes in mode characteristics for any of the modes would not affect the total transportation demand. This is reasonable for marginal changes. Second, it assumes that transportation demand and transportation mode characteristics do not affect long-run choices such as locational choices and activity patterns, which is reasonable for a short-run forecast. (For long-run modelling, activity shifts reflected in a series of feedback loops was a classic but rarely implemented extension.) Third, a fundamental issue with this model is, like many models, the model is only as good as the analyst. In essence, this model takes a bottom-up approach, adding variables that the analyst believe that will be relevant one at a time. While the advantage of this model is that each and every step can be empirically shown to be relevant and robust, the disadvantage is that the lack of empirical data can cause important key variables to be discarded or not taken into account. For example, in Small and Winston (1999), they explicitly recognize that other independent variables, such as whether the vehicle in question is electrically powered, and the amount of luggage space available, can have a measurable effect on the mode choice. In the Manheim model, the effect of these variables would have been captured in the "idiosyncratic preferences constant", thus the model would not have been sensitive to changes in the service level in these variables. While this classic framework is fully extensible to cover any variables that a present-day analyst believe is important and influence travel choices, the classic framework does not provide a way to systematically disaggregate travel choices into its constituent drivers.

2.3 The Planner's Four-Step Model

The Planner's Four-Step Model was developed in the 1950s as a methodology for urban highway planning in the cities of Chicago and Houston (Mitchell and Rapkin, 1954). The basic idea is to forecast the highway capacity required given geospatial data on expected land-use pattern, such as population density per square mile, and other demographics data, such as the number of jobs, households, auto ownership, and other exogenous variables. Although referred to as a Four-Step Model, it is more of a framework consisting of a series of steps which different models could be used in succession to translate the base dataset into a set of transportation flows and thus capacity requirements. The Four Steps are: (1) Generation, (2) Distribution, (3) Mode Split, (4) Route Assignment. (Mcnally, 2000)

Generation refers to the translation of demographic and land-use data into number of transportation generated per unit area. Generally, the trip production is considered to be a function of population and perhaps other variables such as car ownership rates and average household income. This is thought of as the transportation base demand. Conversely, the trip attractions could be calculated based on a function of number of jobs in a given unit area and the type of economic activity that takes place there. Distribution refers to connecting the origins to the destinations. In general, a friction factor is used to determine how far people will travel in order to conduct their desired activity (e.g. employment, school, shopping), and the origins are linked to the destinations. The gravity model is a method often used, which basically assumes that the attraction between an origin and a destination will decrease with the square of the separation distance, similar to the formula for calculating gravitational attraction. Given the demands between a number of origins and destinations pairs, these flows are then distributed over a number of modes and a number of routes. Usually a method similar to that described in Section 3.2 is used for assigning the flows to different modes and routes.

The classic criticism of this type of model fall into a number of categories: (1) the Planner's Four Step Model is designed to calculate traffic volumes given land use, and does not explicitly account for the interaction between transportation infrastructure provision and land-use patterns; (2) the model does not take into account of induced demand, the phenomenon that if the generalized costs of transportation is lowered, more transportation demand would be generated as a result; (3) the mode and route choice portion of the model is based on a discrete choice framework, as a result the model is very mode-based, requiring definition of distinct modes. While hybrid vehicles are still a rarity, it is not clear that the model accurately captures the effect of subtle changes in levels-of-service. In a wide transit network, it may be possible to travel between two given points via a variety of routes and mode combinations. Access can be achieved by walking or with the private auto; some nodes offer rail, bus rapid, express bus, and local bus services, while other nodes offer local bus service only. It is not clear how such subtle effects could be captured in this model, even with a nested-logit implementation of the combined Mode-Split/Assignment stage. While in urban transportation applications, since public transit captures such a small market share of all urban trips, these transit-auto interactions may not be important. However, in mid-distance intercity travel (about 200~600 miles), where truly contestable markets exist and all line-haul modes (air, rail, and highway) capture a respectable share of the market, such micro-interactions could become the driving factor behind the mode choice.

2.4 Discrete Choice Methodologies

Ben-Akiva and Lerman (1985) offers a very theoretical treatment of discrete choice modelling methodologies in their book, discussing the advancement of the field into multidimensional choice, nested logit models. These are extensions of the basic discrete choice framework that were discussed in Manheim's treatise. The focus of the book, however, appears to be computationally estimating such models. Adaptation of the mathematical techniques into a form that will support transportation analysis is left up to the reader. Ben-Akiva and Lerman discuss some of the issues that arise when their approach is adapted for urban transportation planning, but this was not the main focus of the book. There was no explicit mention of application of this methodology for intercity transportation demand forecasting in the book. While it is evident that the approach could be adapted, clearly further work is required in that area.

Where the authors have provided examples of how to apply their model to an actual situation, great care has clearly been taken to ensure that the model is indeed sensitive to what the they were testing. For instance, in their study on urban transportation forecasting, Ben-Akiva and Lerman recognized that disaggregation of originating demands and flows down to a traffic analysis zone level is important. The question a policymaker or a manager is likely to ask may involve decisions to close a street, construct a bridge, or reroute a bus. Much of the impact of these decisions may be felt within a town or a neighbourhood, thus a town-by-town analysis may not capture all the expected congestion effects, or localized ridership changes. They did not provide a specific framework or checklist to ensure the modelling results are valid for the intended purposes. It was generally felt this was the task of the modeller, and not the developer of the modelling methodology.

As alluded to earlier, there is a fundamental problem with this statistical approach, especially when applied by analysts thinking too narrowly. When attempting to calibrate the model, the analysts collect the dependent variable (i.e. mode choice) and collect data on a number of independent variables that the analyst believe will influence the dependent variable. However, if a key independent variable was not believed to be important (or indeed, happen to have very similar values in all the data points that the analyst happen to collect), a model of high statistical significance could be obtained without considering the effect of the key variable. In that sense the model can only be as good as the data that is collected, and the modelling process does not challenge the analyst to seek further explanations of travel behaviour beyond what is exhibited in the dataset.

An analogy in the field of transit performance measures are the approaches presented by Lee (1989) and Fielding (1987), as summarized in Wilson (2001). Fielding suggested a bottom-up statistical approach where a number of performance measures are collected and measures that do not yield much information are progressively removed using correlation and factor analyses. The main problem with this approach is that measures that are not in the initial dataset would not be present in the final results. Lee suggested a top-down approach, disaggregating a cost-effectiveness measure into many constituent measures and searching for representative measures systematically down a tree. Perhaps a better approach to explaining travel behaviour is to start at the top and disaggregate the ultimate mode-choice decision into a series of factors, then collecting data on these individual factors to construct the ultimate model.

2.5 Demand Models Sensitive to Operating Plan

The sensitivity of travellers to the departure and arrival times has long been known. Especially in the intercity sector, there have been attempts to quantify the effects in a model. Slagmolen (1980) examined the concept of "adjustment time" in a study on intercity travel demand. Adjustment time is the extra time added to the trip because schedules do not conform exactly to the travellers' needs. Rather than simply considering service frequency and/or expected wait time, with Slagmolen's model, it is possible to input the entire operating plan for a passenger railroad and forecast the effect of operating plan changes, such as insertion of additional stops into express trains, or scheduling a late-night departure to leave 30 minutes later to capture extra passengers. The model is then making an explicit trade-off between the additional wait time for the passenger wishing to travel at, say half-past-twenty-three, versus those who would otherwise miss the train because they need to leave at midnight.

The classic criticisms of this type of model is that while it does attempt to capture an extra attribute not traditional considered, the model is very data-hungry and the preference parameters may not be fully transferable. It is conceivable that the value of adjustment time depends on the extent to which the trip is plannable. The plannability of the trip would in turn depend on factors such as the need for planning when travelling by a competitive mode, or historically the level of transportation services provided to the region. For instance, in a rural region where even a automobile trip requires substantial planning (e.g. fuel, maintenance, weather, and time-of-day concerns), the value of adjustment time for a highly reliable mode such as rail may be substantially lower than in an urban region with good infrastructure, where the automobile trip can be obtained on-demand. Conceivably the disutility of planning requirements may also correlate with extraneous factors such as the gross regional product or the automobile ownership. Thus, due care is needed when applying this type of models to an environment for which it is not originally designed. In essence, the model is very sensitive to the operating plan, and has been an invaluable aid to carriers evaluating incremental operating plan changes, but it is not a type of model suitable for strategic planning.

Almost certainly independently, Boeing developed a "Decision Window Model" (1996) intended to forecast the effect of multiple-carrier airline operating plans on the consumer. The basic hypothesis is that there is a latent airline travel demand for a given airport-to-airport origin-destination pair, which is dependent on the time of day (i.e. more people like to travel at noon than at midnight). It is then up to the carriers to "cover" that demand by spreading flights out throughout the day that capture the maximum number of riders. Each passenger has a "decision window" which extend for *x*-number of hours around his or her intended departure time. If only one flight is available within this decision window, then he or she is captive to this flight. If more than one flight is available, then he or she could choose based on carrier, adjustment time or other criteria. If no flights are available within this decision window, then the traveller would elect to cancel the trip. Thus, two carriers scheduling their single daily flights both at 8am would capture only half the passengers that they would if the same two carriers scheduled their daily flights at 8am and 8pm respectively.

These effects are certainly replicated in the real world. Carriers have slowly been staggering their departure times to attain better coverage of the market. For instance, the two shuttle operators in the Northeast Corridor operate hourly shuttles, with one carrier departing on the hour, the other departing on the half-hour. The Decision Window Model (DWM) also correctly replicates the effect of "red-eye" flights, where a mini-peak in demand is observed for a flight lasting more than about four hours long, at around 9pm each evening. For international flights, DWM correctly replicate the effect of time-zone changes on customer preference for flights. For instance, on transatlantic flights, most of the demand from London occurs in the morning and early afternoon, to reach New York in the afternoon or early evening (local time). In the reverse direction, most of the demand from New York occurs in the late evening, arriving in London early morning (local time). Of course, this also happens to allow a very simple aircraft-cycle every 24-hours, achieving high aircraft utilization (approximately 16 flying-hours per day).

2.6 Demand Models Recognizing Trip-Chaining

The issue of trip-chaining has already been considered by demand modellers, at least in the urban transportation sphere. Ben-Akiva and Bowman (1995) carried out research in the Boston area and constructed models that represented the entire day's activity, with data based on diary surveys. The model is thus capable of considering trips that are avoided due to either trip-chaining or substitute activities, such as eating at home instead of eating out. In a follow-up paper, Ben-Akiva et al (1996) considered other elements of complexity such as activity time allocation, temporal variation of feasible activities over the day, and distribution of levels-of-service during the day (such as transit frequency, road congestion). Also, in an innovative step, "no-travel" options (tele-commuting, tele-shopping) were explicitly considered, along with information which can cause changes in departure time, mode, and route choice. This work tacitly acknowledge that the simplified approach of the 1970s and 1980s is insufficient to model the complex decisions facing the travellers with many more options. However, the work also acknowledges that even a typical person's daily activities cannot be modelled at a microscopic level of detail, due to the sheer number of permutations possible. In particular, Bowman estimated that there are 10^{16} different possible ways that a typical person with 10 daily activities may structure his/her day, therefore producing different permutations of transportation requirements.

2.7 Airport Ground Access

Coogan (2000) reviewed the current status of public transportation services to large airports round the world (Chs.2 & 4) and discussed a market research approach to planning public transportation service to airports (Ch.3). He postulates that airport ground access market can be divided into two segments: air travellers, and airport employees. Coogan then further classified passengers by geographic distribution and by trip purpose. The 1996 Logan Survey demonstrated that 36% of all traffic (or 13,644 passengers daily) to the airport originated from more than 16 miles away. The 1995 American Travel Survey suggests an even higher figure, 55%, for New England airports. Coogan further proposes a methodology for developing a market research study, in which he suggested that market research for airport access should focus on data such as residence location, trip purpose, destination airport, access mode, origin of access leg, and other like variables. The uses for such a market research exercise includes: developing public transportation schedule, identifying suitable types of access services, and locating access boarding stops.

The Coogan report is titled "Improving Public Transportation Access to Large Airports". It is clear that, within the context of this work, the airport is seen as a trip attractor, a target, a destination in itself. The market research focuses on the destination "airport", and assumes that the objective of the exercise is getting the traveller from wherever he is travelling from locally, to the airport, as quickly and comfortably as possible. No consideration is given to the possibility that the traveller may by-pass the airport altogether by choosing either highway- or rail-based intercity transportation. The report similarly assumes that there are no airside capacity problems, and as long as mass-transit can deliver people to the airport, endless streams of aircraft will come in and out seamlessly to gobble up the same passengers. Thus, the report isn't really what could be termed a regional intercity transportation strategy. At best, it is a strategy to improving airport access in the region. However, improving airport access may not be the most germane objective function if what we're trying to do is to improve intercity transportation between every point in a region to every point outside this region.

The airport focus aside, Coogan makes some good points about local distribution, and integration of the airport into the regional transportation system. Successful rail systems around the world are discussed in the report (Coogan, Ch.5), including a description of the Heathrow Rail Strategy which proposes local services from Heathrow stopping at key stations on the national rail network around London: Ealing Broadway, Wembley Central, Watford Junction, King's Cross, and following the Thameslink alignment to Gatwick Airport and Brighton. Coogan observed that the existing Heathrow Express service is a good substitute for a hackney carriage to a downtown location, but is not useful for other locations in the metropolitan area. London has a strong downtown, thus a downtown-centric approach to metropolitan distribution may make sense, but even with London Heathrow, a substantial number of patrons come from the greater metropolitan area, and expanding service beyond one single downtown node will most likely benefit more people than a single downtown dispersal point, as Coogan observes: "Examination of total trip times shows... only the stations immediately adjacent to Paddington show a time advantage for the Heathrow Express" (p.85).

Although the report states that the analysis is based on unweighted transfer times, the analysis demonstrates an important insight: the door-to-door time is more important than in-vehicle time for the express portion of the trip. "The data reveals... the comparative travel time on a door-to-door basis seems to influence choice." The report is clearly understanding of the customer's needs to reach a diverse range of destinations within the metropolis: "Travel-time characteristics to downtown may not be a surrogate for travel-time to actual destination." (p.87)

The report does not acknowledge that the same concept could be applied to not just the local-access portion of the intercity trip but the entire intercity itinerary. Having acknowledged that an airport-to-downtown airport-express train is not the complete answer, it does not acknowledge that an airport-to-airport, interairport express airbourne vehicle (i.e. an aeroplane) is also not the complete answer. At least for shorter trips, the airport access time and flight time could easily exceed the total journey time by high speed rail, if high speed rail became more easily accessible. Coogan came to exactly that conclusion on a different level -- the connexions to the regional transportation system and better accessibility of the London Underground is in fact a business advantage for the non-express transportation mode, and that convenience factors may drive mode choice more than the total trip times. Although the report is predominantly about airport access, many of Coogan's observations (Ch.5) are applicable to the intercity travel market in general. The Hong Kong Case Study (p.86) suggested that the directness of the service plays an important role in mode choice, implicitly suggesting the transfer penalty is substantial even with dedicated and well-timed connexions.

2.8 Air Travel Demand Forecasting

Much has been written on the subject of air travel demand forecasting, especially by Belobaba who leads the Passenger Origin Destination Simulator (PODS) research effort at MIT and Boeing (e.g. Belobaba (1989), Belobaba & Weatherford (1996), Belobaba & Hopperstad (1999), amongst others). Traditionally, trending on a flight-leg basis had been used by airlines to understand travel demand, and to assign aircraft capacity. Before the deregulation of the American airline industry in 1978, the airlines only had to worry about individual flight legs, interline transfers were commonplace, and the focus is on capacity provision rather than management. Trend-based models, using seasonally-sensitive data, are often used to predict loads. Essentially, PODS expands the standard model to take into account of changes that have occurred since deregulation, and takes advantage of the increased computing power which had become available since then. Significantly, instead of dealing with a single flight-leg, the PODS model is able to forecast passenger demand on an origin-destination basis, with a variety of routings and connecting options. This allowed the airline industry to investigate network effects, and for the first time evaluate the system impact of route changes and hub connection dynamics quantitatively.

Another interesting feature of the PODS model is its ability to simulate the consumer's reaction to the revenue management system. The basic idea behind the revenue management system is that consumers will "bid" for airline seats with an internal willingness-to-pay. In a capacity constrained situation, the bid price would increase with time, as passenger who must travel would be afraid of not being able to secure a seat; where there is overcapacity, the bid price would decrease with time, as the airline would be keen to "sell-off" any unsold seats close to the time of departure. However, for any given flight, the load situation may change from one of overcapacity to one of capacity constraint – depending in real time on the bookings received. The same airplane also services a number of market segments: commuters who must depart early in the morning and return in the evening, business travellers who must travel on specific days or at specific times, leisure travellers who may be able flexible on their travel dates. Thus, for each individual, there is a different willingness-to-pay, a different elasticity with respect to time-of-day and day-of-week, etc. The revenue management system is designed to extract the most consumer surplus from each individual consumers depending on their willingness-to-pay and flexibility to travel at off-peak times.

Belobaba's work has focused on both forecasting the loads and pricing the flights in such a way that the loads are even. In a way, the they are two different facets of the same problem – in operations planning, changes in aircraft rotations occur about once every two months, when the goal is to maximize the productivity of the assets by ensuring that demands can be met with the appropriate gauge of the aircraft (i.e. minimizing any passengers "spilled" due to inadequate capacity); in pricing, the goal is to maximize the revenue given the aircraft you have already assigned to the flight-leg (thus, in a capacity constrained situation, you would hope to capture the passengers with the highest willingness-to-pay). In a modern operations planning model, the two phases would be combined into the same optimization model to maximize profit: the challenge is to assign aircrafts such that the highest profit is realized systemwide. (As you change aircraft assignment, the costs also change as aircrafts of different gauge are flown different distances). Readers interested in this subject, and related subject areas may refer to Lohatepanont (2002) and TRB E-Circular No. EC-040 (2002). Relevant courses taught at MIT do not reference any textbooks (Belobaba, 2002), suggesting a gap in the current teaching literature.

2.9 Explicit Comparison of Amenities, Journey Time and Fares

Using adaptive conjoint analysis trade-off exercises, Spitz & Adler (2003) asked a sample of commuter rail riders in New Jersey to explicitly compare costs of tickets to the level of amenity provided. The customer was asked questions like: "Which option would you prefer? 10-trip cartel for $52.50 and vinyl seating, or 10-trip cartel for $50.00 and cloth seating?" Rather than a simple yes/no answer, Spitz & Adler asked the customers to rate the importance of their preference. The Spitz study went into utility analyses in some level of detail: customers were asked to trade-off across different types of seating arrangements (e.g. stand in aisle v.s. sit in aisle or window of three-across with center seat unoccupied), and reported that the latter arrangement is worth $0.95 more than the former for a Newark Penn-New York Penn trip. The importance scores also revealed something quite interesting: four more trains per hour is worth almost as much as a 20% change in fare or better seats. (Better seats was quite well defined in this study – Spitz allowed consumers a selection of six types of seats from commuter rail seat vendors, and had samples for the participants to sit in prior to completing the questionnaire). More importantly, an 8-minute reduction in travel time is only worth half as much as the better seats.

The result from the frequency v.s. journey time trade-off is broadly consistent with previous studies: 4 trains/hr frequency change on an existing service headway of every 30 mins is equivalent to lowering the combination of adjustment time and expected waiting time from about 15 mins to about 5 mins – in other words, 10 mins time saving; if we assume from previous literature that access time is worth about one and a half times as much as in-vehicle time, it would translate to 15 minutes of effective in-vehicle time saving. The Spitz result indicated that the frequency change was worth approximately twice the 8-minutes reduction in travel time; we calculated that the change in frequency is equivalent to about 15 minutes in-vehicle time saving.

2.10 Performance Based Technology Scanning

In an earlier working paper (Martland et al.,WP 2002-02), Martland hypothesised that intercity demand modellers are not disaggregating variables that drive travel demand in sufficient detail to capture the nuances of the different technologies. If the objective is to evaluate different technologies to find the best technology for a given intercity corridor, it would be necessary to understand the consumer's response to the all aspects of the technology, not simply the new technology's impact on traditional factors like journey time, reliability, and access time.

This review has found that many aspects of intercity travel technology have already been studied in great detail in the current or even older literature. Some of these models only apply tangentially to technology – for instance, the Belobaba model that simulate the interaction between operating plans, demand, and revenue management algorithms, may appear to have nothing to do with technology. However, in the real world, the costs of a new technology may constrain the operating plans to such an extent that it would be necessary to understand the impact of such a new operating plan on demand and revenue before it is possible to choose between technologies. However, all the models reviewed – and probably many others that were not reviewed – are important in evaluation of technologies. The current approach, of using standard utility analysis to understand the customer's reaction to different aspects of transportation operations and technologies, is termed Performance-Based Technology Scanning (PBTS).

In this chapter, a distinct number of methodologies traditionally considered separate domains were reviewed: (1) total logistics cost framework, usually found in the freight demand modelling and shipping decision support literature; (2) classic demand models driven by values-of-time, usually comprising of in-vehicle and access time, plus other factors such as service frequency; (3) explicit willingness-to-pay surveys based on certain amenities or aspects of design for a particular journey. These approaches are all germaine to evaluating intercity transportation system technology and design, and should all be used together in a PBTS study.

The problem of adapting these models to evaluation of intercity transportation solutions has not yet been solved. In the absense of a guiding framework – a checklist of all likely design considerations that will affect customer utility when using the system, it is very tempting to simply adapt a model that has already been calibrated based on a narrow subset of variables that happen to be the most significant factors in a particular situation, and use the model to do strategic analysis. Doing so results in strategic plans that may maximize or minimize a single variable (such as speed) while ignoring other aspects of the design, simply because the design variable were not included in the original model.

By encouraging practitioners to consider total customer utility, it is hoped that in the system design stage, more detailed consideration to design variables would be possible than with a standard approach that simply seeks to minimize journey time between two fixed points with several different technologies. In the rest of this thesis, several cases utilizing the PBTS framework would be demonstrated, showing different circumstances where variables other than average speed could have a dramatic impact on passenger experience – thus revenue, ridership.

References

Barry J. J., Newhouser, R., Rahbee, A. and Sayeda, S. Origin and Destination Estimation in New York City using Automated Fare System Data, *Proceedings of the 2001 TRB Planning Applications Conference*, Corpus Christi, Tex. (2001)

Ben-Akiva, Moshe and Lerman, Steven R. *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass. (1985)

Ben-Akiva, Moshe E. and Bowman, J.L. Activity Based Disaggregate Travel Demand Model System with Daily Activity Schedules, *Transportation Research Records* (1995).

Ben-Akiva, Moshe E., Bowman, J.L., and Gopinath, D. Travel Demand Model System for the Information Era, *Transportation* **23:**241-266. (1996).

Belobaba, Peter P. Airline Yield Management: An Overview of Seat Inventory Control, *Transportation Science*, **21**, 63-73. (1987).

Belobaba Peter P. Application of a Probabilistic Decision Model to Airline Seat Inventory Control, *Operations Research*, **37,** 183-197. (1989).

Belobaba, Peter P. and Weatherford, L. Comparing Decision Rules that Incorporate Customer Diversion in Perishable Asset Revenue Management Situations, *Decision Science*, Vol. **27**, No.2, 343-363. (1996).

Belobaba, Peter P. and Hopperstad, Craig. Boeing/MIT Simulation Study: PODS Results Update, *AGIFORS Reservations and Yield Management Study Group Symposium*, April 27-30, London. (1999)

Coogan, Matthew A. Improving Public Transportation Access to Large Airports, TCRP Report 62, National Academy Press, Washington, D.C. (2000)

Franchisegator. Franchisegator is a franchises web portal which offers much information about franchising in the foodservice and other industries. They are at *http://www.franchisegator.com/cgi-bin/type.php?f_type=11*, Accessed 12 February, 2003. Currently they are listing franchise rates of around 3~5% of gross revenues for the majority of franchises.

Furth, Peter G. and Wilson, Nigel H.M. Setting Frequencies on Bus Routes: Theory and Practice. *Transportation Research Record 818*, 1-7. (1981).

Green, James H. The Irwin Handbook of Telecommunications, McGraw Hill, New York, N.Y. (2000).

Gomez-Ibanez, Jose.(?) Dulles Toll Road Extension Case, *Harvard Kennedy School of Government Case Program*, Cambridge, Mass. (1996?)

Homa, Kenneth E. and Ernst, Ricardo. Benihana Case Material, Operations Management class at the McDonough School of Business, Georgetown University, Washington, D.C. (2003)
At *http://www.msb.edu/faculty/homak/kensweb/POM/Class%202/Class2.ppt*, Accessed February 12, 2003.

Helper, Megan. Tosci's Dead, and I Don't Feel so Good Myself. *Graduate Student News*, February Edition, Massachusetts Institute of Technology, Cambridge, Mass. (2003)

Horowitz, Alan J. and Thompson, Nick A. Evaluation of Intermodal Passenger Transfer Facilities, FHWA Report No. DOT-T-95-02, U.S. DOT, Washington, D.C. (1994).

Hopperstad, Craig A. A Full-Scale Version of the Boeing Decision Window Model, presented to the INFORMS Montreal Conference, Seattle, Wash. (1998). Listed at *http://www.informs.org/Conf/Montreal98/TALKS/TA29.html*.

Industry Insider. Passengers will Pay for Better Service Quality on the Railway, *RAIL* No. **453**, p.32, emap Consumer Magazines, Peterborough, England. (2003).

Lohatepanont, Manoj. Airline Fleet Assignment and Schedule Design: Integrated Models and Algorithms, MIT Thesis, Cambridge, Mass. (2002)

LaBelle, Sarah J. and Fleishman, Daniel. Common Issues in Fare Structure Design, Federal Transit Authoriuty Library, Washington, D.C. (1995).
http://www.fta.dot.gov/library/technology/symops/LABELLE.htm, Accessed February 3, 2003.

Lang, Shef. Taxi 2000 Concept, Fridley, Minn (2000).
http://www.taxi2000.com/, Accessed February 8, 2003.

McClelland, Alison and Macdonald, Fiona. The Social Consequences of Unemployment, The Business Council of Australia, St. Laurence, Australia (1998).
At *http://www.bsl.org.au/pdfs/social.pdf*, Accessed February 12, 2003.

Mohs, James N. Minutes from Meeting of May 2, Connecticut Rail Commuter Council, New Haven, Conn. (2002). At *http://www.trainweb.org/ct/minutes0502.htm*, Accessed February 12, 2003. There are other discussions on the internet, such as *http://talk.nycsubway.org/cgi-bin/subtalk.cgi?read=444525*.

Murga, Mikel. Demand Estimates for a New Toll Tunnel, Bilbao, Spain, *Proceedings of the ITE 68th Annual Meeting*, Toronto, Ont. (1998)

Ortuzar, J. de D. and Willumsen, L.G. *Modelling Transport*, Wiley and Sons, New York, N.Y. (1990)

Psaraki, V., Stathopoulos, A. and Abakoumkin, C. Parking Capacity Requirements for Relocated Airports: The New Athens International Airport, submitted for Transportation Research Records (2003). *http://www.portseattle.org/business/trb/athens2002.htm*, Accessed February 1, 2003.

Parris, Matthew. It's Big, it's Red and it's Free. And it will Save London, Opinion Section February 8, *The Times*, London, England. (2003)
Manheim, Marvin L. Ch.2: The Demand for Transportation, in *Fundamentals of Transportation System Analysis*, MIT Press, Cambridge, Mass. (1979)

Rahbee, Adam and Czerwinski, David. Using Entry-Only Automatic Fare Collection Data to Estimate Rail Transit Passenger Flows at CTA, *Proceedings of the 2002 Transport Chicago Conference*, Chicago, Ill. (2002)

Sheffi, Yossi, Eskandari, B. and Koutsopoulos, H. N. Transportation Mode Choice Based on Total Logistics Cost, *Journal of Business Logistics*, **9** (2), pp. 137-154. (1988)

Small, Kenneth A. and Winston, Clifford. Ch.2: The Demand for Transportation: Models and Applications, in *Essays in Transportation Economics and Policy*, Brookings Institution, Washington, D.C. (1999)

Spicker, Paul. An Introduction to Social Policy, Course Material from *The Centre for Public Policy and Management*, Robert Gordon University, Aberdeen, Scotland (2000).
Specific Citation: *http://www2.rgu.ac.uk/publicpolicy/introduction/needf.htm*, Accessed February 12, 2003.
Publication List: *http://www2.rgu.ac.uk/publicpolicy/introduction/paulspicker.htm*

TRB Committee on Airport Access.
http://www.portseattle.org/business/trb/index.html

Troche, Gerhard. Efficient Night-Train Traffic: Problems and Prospects. Royal Institute of Technology, Department of Infrastructure and Planning, Working Paper No. TRITA-IP AR 99-77, Stockholm, Sweden (1999).

Chapter 3

The Disutility of Time in Transportation

This chapter postulates a new framework for examining the disutility of time in transportation. Traditionally, transportation demand modellers have assumed that journey time is onerous and generates a disutility which is added to the generalized costs of making the trip. The value-of-time is often assumed to be a percentage of patrons' salary. Service amenity improvements such as food service, seating, and entertainment were seen as ways to lower the average values-of-time and thus increasing ridership. Relative effectiveness of improvements is often gauged through either stated preference surveys of existing passengers, or revealed preference data after service improvements were implemented.

There are a number of pitfalls to this approach: (1) the analysts using this approach have not attempted to understand theoretically the effects of new technologies and amenities on the value-of-time, instead are treating it as an empirical result making prediction and sensitivity analyses difficult; (2) the traditional disutility analyses assumes that the time invested in travelling has no other purpose than travelling when in reality, especially in intercity markets, many things with positive utility could be accomplished while travelling; (3) the conventional framework has meant that transportation systems have been designed with users of average values-of-time in mind, when market segmentation (as apposed to aggregation) may achieve higher utility and ultimately revenues; (4) inappropriate uses of values-of-time has often resulted in a transportation system design question being treated as an optimization problem to minimize travel time or generalized costs when the objective ought to be the maximization of passenger utility; there may be multiple combinations of travel time, level of amenities, and other transportation system attributes that gives rise to the same passenger utility when subjected to an out-of-pocket cost constraint.

In this chapter, we develop a framework to enable system designers to take a fresh look at the value-of-time. The essences of this framework are: (1) the value-of-time is not an average value but a continuously changing function throughout the trip, depending not only on the activities that the traveller is taking part in during that time but also potential activities that are available to the traveller;

emerging technologies can influence that function; (2) travel time can "disappear" if the traveller is engaged in activities that would have needed to take place anyway had the travel not occurred; such activities may include sleeping, eating, or simply relaxing; (3) when evaluating values-of-time, a probabilistic evaluation is needed – the probability of one market segment taking part in a certain activity is p[a] and generates a value-of-time for that segment of v(a), thus the design calls for features that will maximize the sum of all p[a] times v[a], not simply the features that maximize v[a]. (3) when making travel choices, individuals are making a constrained optimization decision – given the out-of-pocket cash resources available, how can an individual best leverage that resource to produce the maximum utility associated with a trip (or indeed no trip at all).

3.1 Hypothesis Regarding the Value of Time

Martland et al (2002) postulated that traditional utility models do not evaluate the value-of-time in sufficient detail to examine specifics of technology change or operating change. As demonstrated in the literature review (Chapter 2), carriers have been developing models that forecast the effect of operating plan changes for at least thirty years (Roberts, 1971; Martland, 1972). However, even the more recent models developed in the airline industry (Belobaba, 1996) only take into account of a small number of service attributes. These models are typically highly sensitive at a detailed level to price, cost, and technical attributes such as trip time, number of transfers, frequency and departure time of service. When applied to passenger transportation, these models are often unsuitable to strategic analysis because they are not sensitive to some service attributes that the customer may consider important, such as amenities and access time. Some of the more sophisticated models attempt to estimate the value-of-time based on some function of the prevailing wage, whether this time is spent in the terminal, on board, or accessing the terminal.

The main hypothesis in this paper is that to equate the value-of-time to a function of the prevailing wage rate is misleading, and the value-of-time will depend on how that time is spent, or how else that time could be spent. The amenities available, the physical surroundings, and activity that the consumers are engaged in, will all affect their perception of values of that time. The New Jersey Transit bilevel study examined the impacts of level of crowding, type of seat, and the seat material on commuters' value of in-vehicle time (Spitz & Adler, 2003), supporting our hypothesis. Spitz & Adler demonstrated that the value of in-vehicle time differed by as much as $2.20, of which $1.30 is attributable to a capacity upgrade resulting in less standees, $0.65 is due to the type of seating and $0.15 is due to the type of luggage rack. Such empirical work is valuable. The transportation community have been aware for a long time that by adding concessions or providing climate controls at the terminals, travellers' average disutility-of-time at the terminal could be reduced. However, without more such empirical work, it is difficult to say how much, and it is difficult to ascertain whether climate control is more important than presence of concessions.

Furthermore, Martland et al (2002) suggested that the value-of-time depends on the motivation for spending time. In general, if the consumer chose to spend that time, the disutility of time tends to be very low; if the consumer feels trapped and feels that they must spend that time, the disutility of time tends to be higher. This paper builds on Martland's work by demonstrating the link between consumers' motivation for investment (in time), consumer's resulting quality of experience, and the disutility-of-time incurred while en-route between origin and destination.

Dealing with the average values-of-time over a large time period is counterintuitive. For instance, using a theatre example, if a two-hour movie cost $8 in admission, it could be argued that the utility gained by the consumer per hour of movie is about $4 per hour. However, consumers are generally unwilling to pay $6 to watch the first 1.5 hours of the movie; on the other hand, latecomers will often pay $8 for the last 1.5 hours of the movie. The entertainment example is appropriate since transportation industry is becoming more of a service industry as user-produced transportation (the automobile) is becoming more widespread. Thus, it is conceivable that the time spent on-board vehicles and in terminals could be further subdivided into different activities, each with different values of time. For instance, the disutility-of-time while standing in line at an airport waiting for check-in is most likely very different from the disutility-of-time after check in when the travellers may browse around in the airport mall.

3.2 How the Value-of-Time Relates to Concessions

Interestingly, although perhaps unsurprisingly, the literature in the economics of airport and transit center concession development has mainly focused on the success of the concession as a going concern rather than the value the retail development is adding to the traveller experience in the terminal. Most literature mention that the presence of concessions add to the traveller experience (Bay Area Economics, 1999); airports in Europe studied the possible effects of abolishment of intra-EU duty-free in 1999 on competitiveness of airports (Hopkins, 1998), competitiveness of international versus domestic vacation destinations, and its expected impact on stores that specialize in duty-free goods (Commission of The European Communities, 1998). There appears to be little quantitative work done specifically on the effect of the concessions on the traveller's experience, although its positive influence is well documented. However, the reverse has been established, at least in one study: The House of Common contended that abolishing duty-free might encourage smuggling, and that the carriers profit from concessions onboard the vehicle or at the terminals (House of Commons, 1999).

This suggests some will travel more frequently or pay more because of the possibility of enjoying the duty-free concessions, and that the carrier is able to leverage additional profits from the consumer surplus resulting from decreased disutility-of-time while on-board due to the use of vehicle-bourne concessions. While the House of Commons report gives us an idea of the sign and the order of the benefit, it is not a quantitative assessment. Decreased travel demand in light of abolishment of duty-free would suggest the elasticity of travel demand with respect to the price of duty free goods is nonzero. If we hypothesize that the reason the elasticity is nonzero is because the purchasing of duty free goods are resulting in a consumer surplus, the nonzero elasticity suggests the surplus is affecting demand for travel and hence the utility of travel, it is reasonable to suggest that by removing the airport concessions, thereby removing the possibility of generating that surplus, travel demand would be aversely affected. The reverse may also be true: by adding concessions where they do not previously exist, travel demand may be stimulated.

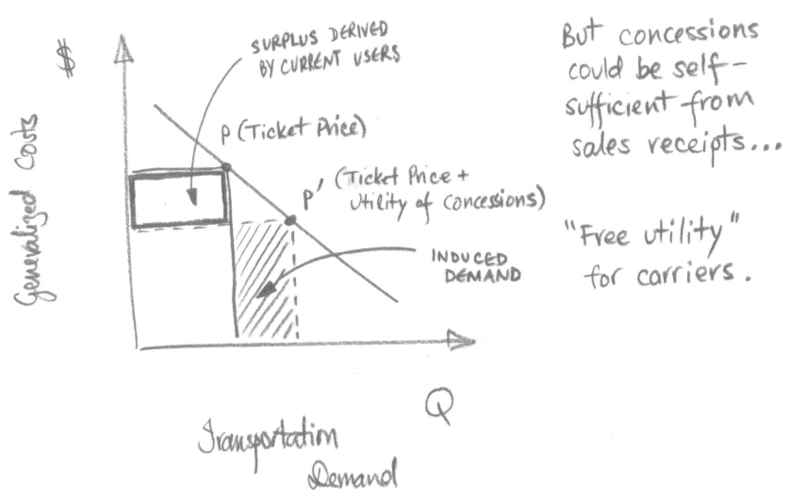


Plate 3.1: Utility is Free

<u>The Channel Tunnel Example</u>

Following the construction of the Channel Tunnel, the crossing of the English Channel with an automobile was cut from a four-hour ferry trip to a two-hour trip onboard a freight train. Originally, the ferry operators thought that their core business was being undercut and asked for government compensation. When the government compensation did not materialize, they chose to refit some old cruise liners to offer luxury overnight service. Although the ridership declined compared to pre-Chunnel levels, the carriers found a clear market niche and continued to operate profitably. Clearly, the onboard amenities were a sufficient attraction to compete at some level, compensating for the longer journey time.

<u>The Fishing Example</u>

When a user is choosing to use a service, the user is making an investment in time to use the service. This is a well-known concept in transportation systems analysis called "generalized costs". However, when economists are evaluating other service industries, they do not generally include explicitly in their framework unless they are explicitly talking about time consumed by transportation as an overall vacation schedule. For instance, two studies in Central Idaho contained quantitative data on the number of vacation trips taken per year versus travel time (Normandeau Associates, 1999). Based on survey data, the study demonstrated that the demand for fishing is sensitive to both the trip cost, and the access time. Significantly, the study confirms what seems intuitively obvious:

> Each angler has a different travel cost (price) for a sportfishing trip from home to the river. Variation among anglers in travel cost from home to sportfishing site (i.e., price variation) [… is significant.] Non-monetary factors, such as available free time and relative enjoyment for sportfishing, will also affect the number of river visits per year. The statistical demand curve should incorporate all the factors which affect the publics' willingness-to-pay for sportfishing at the river.

This suggests that the demand for fishing vacations depend not just on the price of the fishing permit and other costs, but also on the investment in time required to get to the site. The number of fishing trips per year drops significantly beyond about 7.5 hours or $100. These results, backed by survey data and commonsense, suggests that the intercity demand modeller ought to consider a much broader range of variables than in-vehicle time in intercity transportation, especially when vacation travel is involved. The recognition that each angler has a different travel cost is key – in Chapter 5, the idea that each intercity traveller has different access costs to terminals serving different modes will be explored in much greater detail.

Economic research has recognized for a long time that the time taken to search for a service can be considered the cost of using a service (e.g., Stigler, 1961), and that good advertising and product accessibility can lowers that cost to the point that it's not worthwhile for the consumer to invest additional time in searching for a marginally cheaper or better alternative. The lowered communication and search costs have also been implicated in the creation of cities (Mills, 1993). It is conceivable that a conceptual economic model could be formulated that breaks the generalized cost of using a service down to three components: (1) time taken to locate the service; (2) price charged by the service provider; (3) time taken to actually consume the service. The Stigler concept suggests substitutability between (1) and (2). The anecdotal evidence suggests that (2) and (3) are actually also substitutable, as evidenced by various initiatives that are offered by service industries that are queue-intensive: turnpike toll-collection, airport check-in, mail order companies and other retail outlets that charge extra for "express service". All these evidence suggests that the time taken to consume the service could be considered part of the generalized costs of using the service, and it is a necessary investment on the part of consumers choosing to use such service. In the restaurant industry, anecdotal literature suggests the time taken to prepare food at restaurants is a driver in restaurant demand (Foodservice.com, 2002), which would be consistent with the aforementioned economic model. In the next section, we will examine the restaurant model in more detail.

3.3 How the Value-of-Time Relates to Foodservice

It is possible to illustrate this using the example of time spent eating in a restaurant. An investment in time is required to consume a meal in a restaurant, including the need to wait between ordering and serving. If prevailing wage rates were used to evaluate the value of time spent waiting and actually eating, it is not clear that restaurants are in general a very profitable proposition. The restaurant sells a product (a meal) at much more than the cost of the competition (cooking at home), with conceivably a much longer waiting time (access time to restaurant, queue at restaurant, plus time between ordering and serving). Yet restaurants are clearly a viable proposition, even in low-income economies. It is not clear that the amenities offered by the restaurant in fact creates sufficient consumer surplus to overcome the alleged disutility generated by the time-investment required. We postulate that the reason the time spent at a restaurant is grossly undervalued by the consumer because the consumer in fact is choosing to invest that time. Even with the same physical surroundings, if the consumers feel that they have no choice, restaurant economics would simply not work.

Basic economic theory suggests that there is a supply for food, and a demand for food. You will pay for food at a price if the food generates more utility than the price; the difference between the two is then the consumer surplus: the food is worth more to you than what the producer is charging, so you buy it.

3.3.1 The Culinary Utility Model

What does the food do to the value-of-time? Let's examine this in terms of the value-of-time framework. The consumer derives a positive utility from eating food; the consumer then has to spend time eating the food he just bought – an investment in time; the consumer then isn't hungry anymore – hopefully the increased value-of-time while not feeling hungry, discounted back to the point of ingestion, compensates for the investment in time in eating, and investment in out-of-pocket dollars expended to buy food. The "consumer surplus" associated with the food is simply the value of not being hungry discounted back to time of ingestion minus the time spent eating food. If the consumer surplus is negative, then obviously the consumer isn't hungry enough; he will wait another few hours, until hunger-prevention becomes more critical; until the increase in future value-of-time by hunger prevention exceeds the present time-investment required to complete the ingestion process. In equation form:

Consumer Surplus from Food = – (Price of Food + Value of Time Spent Eating)
+ (Utility of Eating + Discounted Hunger-Prevention Value of Time)

Let's call this the Culinary Utility Model. This helps us explain why you hear people say, "I'm too busy to eat now, I'll eat later". The value-of-time before the deadline is reached dominates the disutility of enduring hunger. When faced with expensive food of poor value, sometimes people don't feel hungry; faced with all-you-can-eat buffets, people eat more than they normally do. Really, people in general are quite efficient project evaluation machines. Most people just don't realize it.

3.3.2 The Classical Justification for Making a Trip

Putting this in a transportation context: how do you change the value-of-time of someone either (1) sitting in a terminal waiting for a connexion, or (2) sitting aboard a vehicle waiting to get to the destination? Take the latter case, the justification for sitting in a vehicle is usually because whatever event that is anticipated at the destination (a business meeting, or a vacation) has a high value. People usually trade-off the generalized costs involved in getting there and back against the wonderful utility they will derive by being there. We can null-out the value-of-time "wasted" in getting there if we can somehow give the consumer another justification for investing his or her time. The consumer then wrongly attribute the time spent "in-vehicle" as the time spend doing something else, and the value-of-time in-vehicle decreases to an insignificant value. Although people might be efficient project-evaluation machines, most people are really quite bad at joint-value attribution.

This is an important concept, and has widespread implication in the world of transportation demand modelling and investment evaluation. Let us illustrate with an example, seen from the traveller's frame of reference:

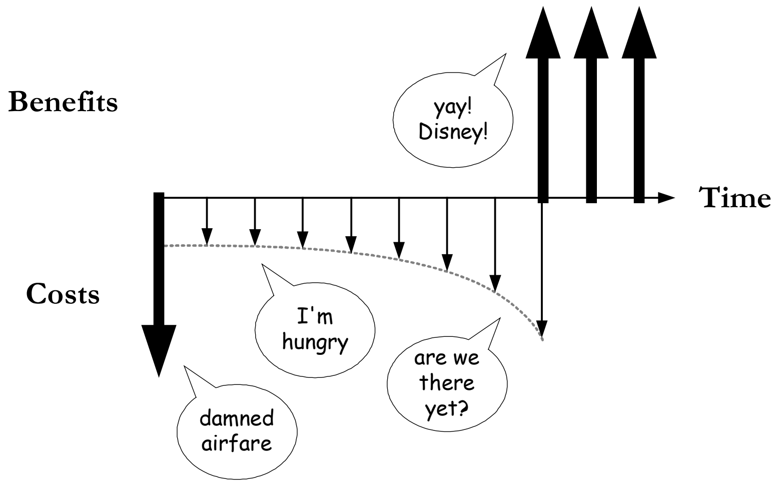


Plate 3.2: Are we there yet?

Exhibit 3.2 illustrates the way in which a traveller may evaluate a trip. The traveller has to make all the investment up-front: out-of-pocket costs of carrier fees, and the investment in time in travelling. We hypothesize that the value-of-time is different for different parts of the trip – in this case, we simplify the trip to a single-segment door-to-door trip in which the traveller takes part in no activities except sit in a seat, we might expect the disutility per unit-time to increase nonlinearly with the time already elapsed. Even with simple trips, the disutility of time is not necessarily a linear function of time. In making a go/no-go decision about the trip, the traveller evaluates the expected benefits of the destination (in terms of utility generated), and compares that to the investment costs required to get there. If the cost/benefit ratio is less than one, the decision is a go. In some cases, as discussed previously, the traveller may have other constraints such as the capital available to invest. Some travellers will use a decision rule; others will use different implicit methodologies of project evaluation (such as payback time analysis, etc.)

3.3.3 Changing the Travellers' Value of Time Through Food

Supposing the carrier now elects to make food service available either on-board or during a particular part of the trip, for example, in the terminal. Assuming that the food is made available at marginal cost, so the travellers purchase socially optimal amount of fude. Thus, anyone who can afford it buys some food, stores it until he or she is hungry, eats it such that hunger is prevented during the time on board. How does the cost-benefit diagram look now?

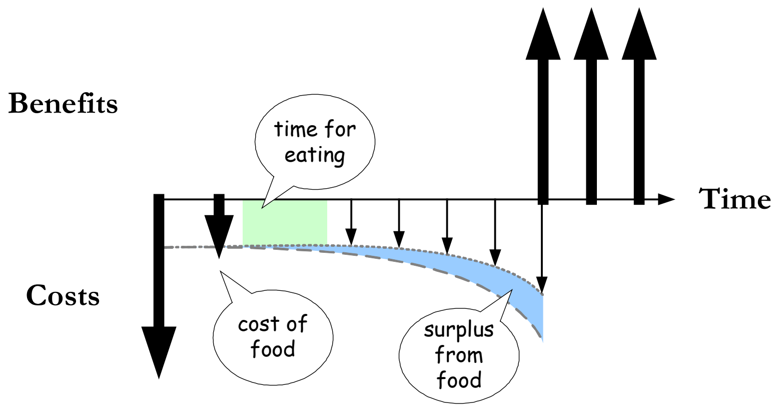


Plate 3.3: Are we there yet? (with food)

Exhibit 3.3 shows the evaluation diagram, from the travellers perspective. The out-of-pocket transportation costs and benefits of taking the trip remain the same, but now there's the incremental costs of buying food. Under what circumstances do people buy food? The Culinary Utility Model tells us that people will not buy food unless the price of food and value-of-time spent eating it is justified by the increase in future value-of-time by hunger prevention. The food costs are therefore attributed entirely to hunger-prevention; the value-of-time spent eating is also attributed to future hunger-prevention.

All this occurs on board a carrier vehicle, during which time the traveller is also moving in space. But we just attributed the time-taken-to-eat to hunger-prevention: a self-sufficient profit center in its own right. So does this mean the same time expended eating on board, during which we are also moving in space, comes free? No longer are we investing time in order to move in space; we are investing time to prevent hunger. From the perspective of the investor who likes to move in space, the time expended eating on board is a sunk cost: "I gotta spend this time eating anyway – I am hungry!" So the cost-of-time involved in hunger-prevention and movement-in-space are *joint costs* which the traveller allocates exclusively to eating. The traveller is reaping economies of scope in joint time-utilization (normal people call this "saving time" or "doing two things at once"). To achieve movement in space, without investing time, is called teleporting. If you believe this time-cost-attribution model, teleporting is really possible.

Exhibit 3.3 also shows another very subtle point. The time between when the traveller finishes eating, and when the traveller arrives, the traveller is sitting on board a vehicle, and investing time to achieve the movement in space. Let's ask this hypothetical traveller a simple question: "Would you rather travel hungry, or having had some food?" Question of motion-sickness aside, most would prefer to part-take in everyday activities, including travel, while their stomach are not making gurgling noises. In other words, the value-of-time in vehicle is lower if the traveller is not hungry; being hungry on board a vehicle adds to the disutility of simply sitting bored. Thus, by making food easily available, you are actually changing the value-of-time on-board a vehicle by encouraging travellers to feed their stomachs.

3.3.4 Issues in Railroad Café Economics

You might argue that the baseline case is a traveller travelling when he or she is not hungry; the fact that he or she hadn't eaten adds to the disutility, and making food available brings it back to the baseline level. Is this correct? Two issues with this argument: (1) Business travellers are too busy to eat before departure; students are too disorganized to eat before departure, so people end up hungry. (2) A carrier with relatively long in-vehicle times and no possibility of en-route stops, such as the high-speed train, would compete with modes such as air which offers short journey times, and modes offering unlimited unplanned stops such as the private auto. People will get hungry. Thus, making food available either at terminals or on board is key to a rail carrier, but not to others, since the other carriers offer other means to combat hunger while in-transit.

Why should food be priced at marginal cost? Pricing at marginal costs mean that everyone who is hungry enough to justify food production will get food. It ensures that the evaluation process for deciding whether to buy food or not is based on entirely the same Culinary Utility Model whether you're in a grocery store, at a terminal, or onboard a vehicle. On board a vehicle or in a terminal, there is always the temptation to price up since the customers are perceived as captives. Yes, they are captives, but if you price at more than the marginal costs, that means *some* passengers who want food will not get food, because they can wait until later – or perhaps take another mode of transportation the next time just so they won't have to go hungry. If the passenger elects to wait, you the carrier have just made his in-vehicle experience miserable. If the passenger elects to travel by another mode to avoid the hunger or the overpriced food, well, you've just lost a ticket revenue. Either way, you can't win, because the customer utility is tied to his or her stomach and to your bottom line.

Why not subsidize food? Not only does it lead to wastage, some passengers who do not need to eat will end up with food just because it's cheap. To the carrier, or the provider of the food, this represents a *deadweight loss*. Of course, any such loss-making operation eventually comes back to haunt the carrier through increased ticket prices to cover increased costs. This will tend to divert customers to other modes which do not have such overhead. After all, if you're in a private auto, you have near perfect competition in the market for food provision, and price is going to be pretty close to marginal costs. The railroad café is not a monopoly; it competes with all other food providers available to all other transportation modes, and it has an important effect on the railroad's ticket revenue through the value of time.

This is not necessarily an argument for the Café Car. A Café Car might be an effective way to sell food, or it might not. The whole point is that the access to food for those who are prepared to pay the marginal costs must be available. The provision of a microwave oven, a hot water source, and a vending machine in a vestibule area may be all that is needed. The facilities that are made available converts onerous on-board time into neutral time spent cooking as if in one's own kitchen. In the context of the maintenance and capital costs of railcars, these facilities are loose change. By providing a Café Car, the carrier is providing a value-added service, an opportunity to dine on board or purchase food services in a café setting. Like the roadside café or diner, this feature should be self-sustaining. Perhaps legislative changes would be required owing to liability concerns before this type of scheme will become practical.

I've caught myself refusing to eat at airports because of the inflated prices of airport concessions and thus making my own journey very uncomfortable. Moms will pack food for college students going off to college; moms will tell you always to travel with food. No one calls a convenience store or a gas station a luxury. A restaurant might be a luxury, but it's providing a service at a price the consumer is willing to pay, and it's making it (i.e. the cost of production is below the price charged). You'd be pretty dismayed if you drove into a small town and didn't find any restaurants or any stores, no Dunkin' Donuts and no McDonald's. That sort of place is called a ghost-town, and people tend to minimize the amount of time they spend there. On the other hand, it's hard not to stop while driving through a quaint village peppered with touristy places. The presence of businesses, even those that charge a little more than marginal costs, clearly enhance the value of time in the close proximity. People like to trade. People go to great lengths to trade – they even created the silk road.

3.3.5 Food Can Make Onerous Journey Time Disappear

Let us complete our discussion on food service with some old railroaders' wisdom. Railroaders (and truckers, too) are kind of a special breed. They always know where to go eat. They always have food, even if it is just a can of spaghetti that explodes in the engine room of a Missouri Pacific locomotive (Santucci, 2001) – because you don't want to be "stuck hungry on a train". Employees are on the system day-in and day-out; employees know that they do not like to be stuck without food. Customers don't necessarily know this, or have the organizational skills, time, or local geographic knowledge to buy food. Let's feed them.

"Journey Time Just Magically Disappears"

I originally coined this phrase with respect to overnight services, meaning that the time spent sleeping is nearly equivalent to the time spent sleeping at home. Thus, if one is able to sleep while achieving movement in space, in effect, the movement in space becomes a byproduct of whatever other activity the traveller chooses to engage in. The onerous "journey time" has then *disappeared* from the traveller's mental model of the trip, because the traveller attributes that time to something else. We have already seen the application of this principle to the time spent eating on board. It also applies to a variety of other activities that could be performed on-board, such as work, entertainment, rest, and socializing.

3.4 The Dollar Value of a Day Methodology

The Dollar Value of a Day Methodology is gaining increasing recognition amongst the academic community. Britain, for example, has conducted surveys in recent years that attempted to calculate the economic impact of household or voluntary work (BBC News, 2002). According to analysis performed on data generated in The UK 2000 Time Use Survey, the average hourly rate for housework is £4.72 (about US$ 7). In Switzerland, a recent paper describes an analysis of the 1997 Swiss Labour Force Survey, where the allocation and value-of-time assigned to housework and childcare was analyzed. The value-of-time quoted was calculated with two market replacement cost methods and three opportunity cost methods (Sousa-Poza, 2001). In Spain, where the food market is being influenced by rapid cultural and economic changes, a study determined that high-income families are more likely to consume food away from home and spend more than others on food away from home (Justo & Jensen, 1998). This tacitly suggests that in effect, different families have different evaluations of the value-of-time required in preparing food.

The British analysis is driven by The UK 2000 Time Use Survey, which is a field-survey involving individual questionnaires and telephone follow-ups. Selected household heads or their partners completed a household questionnaire. All individuals aged eight or over were asked to complete individual questionnaires, two one-day diaries and a one week work and education time sheet (National Statistics Online, 2002). The data provided no information about how these individuals valued the time, but the aggregate results enabled useful insights into how individuals allocate their limited time. For instance, the average Brit spends about 80 minutes each weekday eating, a further 50 minutes each weekday on "food management", but only about 80 minutes per weekday travelling. Significantly, the average Brit spends up to 140 minutes each weekday watching television or video. Interestingly, in an independent study (Schafer, 1999) using American subjects, Schafer concluded that the average person spends about 90 minutes per day travelling, and a faster mode simply induced people to travel further each day (for work or other purposes). The 2000 U.S. Census, which gave an average commute time of 94 minutes, is consistent with Schafer's findings. Given the totally different demographics, and perception of longer commutes in the United States, it is surprising that the time spent travelling is very similar. The data are not directly comparable, since the 80 minutes cited in the British survey included the non-work population who may spend time travelling to a shopping mall, while the Census data deals in work-travel only, but the consistency of the data is striking.

In the U.S., a comparable data source was the Dollar Value of a Day publication available from Expectancy Data (Ward, 2002). Ward and Associates used a similar time-diary data provided by the National Human Activity Pattern Survey (Environmental Protection Agency) and a Department of Labor Survey of wages by occupation by geographical area.

The common element of the Dollar-Value of a Day methodologies reviewed is that they only attempt to evaluate those parts of the day which are actively engaged in producing monetarily valuable work for others. For instance, in the housework category, the recipient of the benefits is everyone in the household, while a specific task might be carried out by a specific member of the household. In that way, they produce external benefits which can be evaluated using the replacement cost method. If the household member was not available to do the specific task, what would be the market rate for hiring a replacement for that length of time? In some other surveys, the opportunity cost method asks the question, if the household members were not engaged in performing that task, what value could they generate (for others) if they performed other tasks? Neither method was able to evaluate the question: what is the value of time for the time that individuals invest for the benefit of themselves?

<u>When Producer and Consumer of Services is the Same Individual</u>

The value of time where the individual is both the producer and consumer of services is a particularly interesting question economically because it is both difficult to estimate by conventional methods and related to many transportation problems. It is difficult to measure since it's not possible to hire an individual to produce the same services – if you hire somebody to eat for you, you don't exactly gain any weight; if you hire somebody to travel for you, you don't get to go on the trip. While it is possible to ask a subordinate to make the trip for you in a business situation, most business travel involves at least some kind of incentive for the individual (whether that be because the individual values the opportunity to attend the conference for personal career development, or the individual just likes to travel), and pleasure travel requires the presence of the individual. The conventional argument that the consumers' willingness-to-pay in out-of-pocket price for the trip represents the utility they derive from that trip is misleading for two reasons: (1) the price is determined by competition in the market, consumer surplus is generated over and above the investment, thus the price is in fact a lower-bound on what the individual is willing to pay for that trip; (2) the individual is investing time to make the trip, and it is not clear how that time should be valued. Conventional methods assume that the individual would otherwise be at work, but this may not necessarily be true. Other leisure options are available, but most individuals choose a mixture of leisure options over the course of one year. While welfare-maximizing economics will suggest that the individual will persistently choose the leisure option that generates the best "return" on the time invested – other than adamant philatists, electronics enthusiasts, and train spotters, it is rare that individuals will devote all their free time to the same leisure activity. There are also other activities which are necessary to sustain life, such as eating. The value-of-time spent eating (or utility derived from investment in out-of-pocket cost and time seeking and eating food) is likely to be somewhat dependent on the level of hunger. The value-of-time in those cases could be difficult to evaluate.

How does this relate to transportation? The case of the automobile operator on a business trip, with a bagel in hand and talking on a personal cellphone while listening to the news on the radio demonstrates that a person can easily be fulfilling lifetime necessities while accomplishing both personal enjoyment and a business goal. Such individuals are not only consumers and producers of transportation services at the same time, but also utilize the same time to take part in activities that generate surplus both personally and perhaps for others.

<u>Marketing-Informed Design of Transportation Systems</u>

In transportation systems design, the system should be designed to facilitate activities for a range of different user groups, and not simply the "average" user. The type of activities that are likely to be of value to travellers could be found through analysis of time-use survey data. The most leveraged activities – in the sense that the type of activities that travellers would most likely to want to take part in during a long journey, are presumably the ones where individuals spend most of their time doing during a day, such as sleeping, watching television, or eating. The Metro newspaper became an astounding success around the world in transit-friendly cities because the Metro's owner realized that he could produce a newspaper that are suitable for reading in between transit stops. At the margin, the Metro probably produces a small benefit for the carrier, as a subway ride without reading material incurs more disutility than one with.

Transportation users are individuals, and it is these differences in individual characteristics that we must exploit to consider design issues. The ideal transportation system may not be the same for every individual, but it is the designer's goal to ensure that it is close to ideal for a sufficient number of people that the investment is justified. Systems that are tailored to the average values tend to end up being unsatisfactory for every individual. It is possible to argue that those who eat or socialize while they travel are the minority or the exception, and for most people the time they spend travelling are wasted and thus must be compensated for by an explicit remuneration at rates similar to their wage. But there are many different alternatives in transportation systems, and the goal of creating alternatives is so that certain alternatives can attract certain types of consumers. By tailoring the system to cater towards certain type of consumers, not only do we do "better than average" in terms of value-of-time, the consumer's behaviour can also be influenced – hence altering the average value of time for a particular mode, market segment, or any other population.

3.5 How the Value-of-Time Relates to Time-Use Surveys

The Time-Use Survey data can be further analyzed to gain insights into consumer behaviour, and how travellers might like to spend their time on board a transportation vehicle for prolonged periods. Intercity transportation, typically associated with door-to-door travel times of at least three hours, will necessarily create a disruption in the day's activities. Research into how transportation companies could leverage the time that consumers spend in their custody, both to generate additional revenue from offering additional services, and to generate better market share by allowing the consumer to utilize the journey time (especially on the slower modes) more effectively.

Interpreting Aggregate Time-Use Survey Data

The UK 2000 Time Use Survey suggests that, amongst the adult population, the average male spends 520 minutes per day working at a paid job outside the home (8.7 hours), and a similar amount of time sleeping. However, the average female only spend 350 minutes per day working (5.8 hours), but 520 minutes sleeping. This does not necessarily mean that females who do work spend less time per day working than their male counterparts. It probably reflects the fact that the percentage of adult females who choose to work is smaller than the percentage of adult males who choose to work. Thus, it is quite difficult to predict individual behaviour (on which individual travel decisions would be made) based on the average time-use data. Nonetheless, this does not mean the aggregate data is unusable; it can in fact be quite useful, if one is careful to notice such aggregation effects. When combined with the shape of the distribution, time-use data can be very helpful indeed.

Understanding How Travel Time Could be Used More Effectively

The average individual spends 80 minutes per day travelling. It is possible to hypothesize that this represents the time taken to commute daily from work, and on non-work days this represent the time taken to travel between leisure activities and maintenance activities (such as shopping). It should be highlighted at this point that much research has been done already in building activity-based models of travel demand (See Chapter 2 for a review). The current approach is much more proactive – given the typical activities that take place during the day, and given that the individual has decided to travel, how can we enabled the "dead" time invested in travel to be used more productively?

The key to integrating the travel experience with an individual's daily routine, and thus making the time "disappear", is to identify the activity which takes up a large portion of the day for many individuals, and enable those to take place while travelling. Some activities would not be practical on board transportation vehicles, but could be practical in the terminals. Other activities are possible on board at least some vehicles. The average individual spends 520 minutes per day sleeping; 220 minutes watching tele; 160 minutes eating; and 50 minutes a day socializing. These are activities which are perfectly practical on board vehicles. By presenting a range of options to occupy the consumer's time that would otherwise be idle and invested solely for the purpose of travelling, the carrier is able to leverage the value of much of that time. Different people will value that time differently, but provided that the carrier is able to provide the option on a break-even basis, it really does not matter what the value of that time is; the value of that time is known to be positive.

3.6 Consumers' Attitudes Towards Travel Time

Previous research in social psychology has demonstrated this phenomenon in quite a different setting. Miller and Form (1951) described an experiment at the Hawthorne Works of the Western Electric Company in Chicago between 1924 and 1927. In this experiment, team of six girls were chosen to determine the production rate of telephone relays against such variables as rest-breaks, availability of free hot-meals at lunchtime, and physical conditions such as lighting. The increased amenities apparently caused an increase in output. However, at the end of the experiment, the team was returned to the same physical conditions, the output increased even more. The explanation offered by Brown (1985) in his review is that by asking for the cooperation from the girls, the experimenter has made them feel important and thus the girls' attitude to the work changed, and they worked harder than ever before, irrespective of the physical conditions. The fact that the experimenter listened to the workforce and pampered them was producing increase in output far greater than the loss resulting from rest-breaks and shorter work hours. Mayo (1933), author of several textbooks on the subject, researched the behaviour of Philadelphia mill workers, reached broadly the same conclusions.

Applying the concept to transportation would suggest that the consumer's attitude towards the time he or she is spending on onboard the vehicle is important. If the transportation company makes amenities or other entertainment options available that signals to the consumer that the company cares, the disutility of time could be reduced or may even become non-existent. This research would suggest that the economics of provision of amenities should not be analysed on a case-by-case basis; it is the existence of options that allow the consumers to spend the time as they please which generates the surplus, and not the existence of any particular amenity.

This would also be consistent with the hereto unexplained dominance of the automobile in medium-distance transportation. Despite the long journey time, the automobile achieves 8.3m trips per year, much more than the 0.2m trips by train (American Travel Survey, 1995) and roughly 0.5m trips by air. Some (Morrison & Winston, 1985) have attempted to explain this in terms of unrealistically low disutility-of-time of over-the-road intercity travellers. Significantly, in the Morrison & Winston study, the data was collected from vacation travellers. These travellers are clearly choosing to spend the time in the vehicle, where the drive itself is part of the vacation experience. The travellers are also able to choose to spend the time as they wish -- vacation highway routes are typically peppered with entertainment options ranging from restaurants, hotels, scenic spots to casinos en-route.

Exactly how much of these options can be replicated by intercity rail carriers economically remains to be seen, since it is both a function of carrier technology and the actual cost of the entertainment business. However, it is conceivable that with creative design of terminals and vehicles, sufficient competitive choices of activities could be offered by long-distance carriers to change the attitude towards the investment in travel time of most travellers. Another observation to follow logically is the importance of educating the consumers on ways of enjoying the journey. The auto industry spends a great deal of its budget on reinforcing the auto dominant image, and stressing the freedom element associated with the automobile. The consumers also respond by spending a great deal of resources on luxuries onboard the vehicle.

3.7 Conclusions

The most important conclusion that follows from this research is perhaps that evaluation for high-speed rail or other intercity transportation technologies should not focus too narrowly on the technical attributes such as journey time, frequency, and capacity. Equally important are the human attributes of how the time on-board could be spent and the ability of the technology to adapt to changing human demands.

In this chapter, we demonstrated that concessions and other activity options en-route influence transportation demand sufficiently significantly that a carrier should take notice. We proposed a model for understanding how people evaluate decisions regarding how to invest their time. Using this model, we were able to explain why dining concessions, shopping opportunities, and other amenities can make journey time 'disappear' – essentially by incurring transportation time as a joint cost with another activity. To determine what kind of amenities are most leveraged, and most likely to be applicable to many market segments, we briefly reviewed results of time-use surveys. Thus, a proactive approach to transportation systems design is introduced – given the typical activities that take place during the day, and given that the individual has decided to travel, how can we enabled the "dead" time invested in travel to be used more productively?

In examining the nuances of technological enhancements of the travelling experience, it is necessary to examine the nuances of the travel experience. If new technologies can enable mobile communication from within the train, we can quantitatively determine what percentage of travellers would derive a higher value of time as a result, and compare that to conventional technologies that may reduce travel time, or service enhancements that would provide food or other amenities. By evaluating the cost and benefits of each of the options, using the value-of-time framework, we are able to rationally prioritize all possible service enhancements or technological innovations. It is likely that a combined package featuring some journey-time reductions, some service enhancements and some hi-tech gizmo would be optimal.

In large metropolitan areas, where further capacity expansion of the highway network is nearly impossible, rail's ability to haul large amount of weight cheaply and its relatively small terminal footprint may become a strong advantage in offering the diversity of entertainment options en-route that the consumers are likely to expect. The ability to haul weight increases the options of entertainment on-board, while those elements that are not available on-board could conceivably be offered at the terminal. The small terminal would enable such options to be offered close to population centers, decreasing the reliance of such businesses exclusively on travellers -- as is the case at an airport mall. In addition, the small terminal footprint may allow multiple terminals to be constructed in an urban area, allowing substitution of onerous access time for more pleasant in-vehicle time – an idea explored further in Chapter 5.

References

BBC News. Housework Worth £100bn Annually, British Broadcasting Corporation, London, England (2000). http://news.bbc.co.uk/1/hi/uk/1948016.stm

Brown, J.A.C. The Social Psychology of Industry, Sidgwick & Jackson, London, England. (1985)

Bay Area Economics. Evaluation and Market Positioning for Transit-Oriented Retail for Arlington County, Virgnia. Berkeley, Calif. (1999).

EU Duty Free Working Paper – http://europa.eu.int/comm/taxation_customs/publications /official_doc/sec/taxation/sec98_01994/sec98_01994en.pdf

Foodservice.com. Fast-casual Restaurants Quickly Expanding Their Niche, Friday October 25, 2002. http://www.foodservice.com/news_homepage_expandtitle_fromhome.cfm?passid=4922

Hopkins, Kelvin (M.P. for Luton, North). Proceedings of the United Kingdom Parliament, May 6, Westminster, England (1998) – http://www.parliament.the-stationery-office.co.uk/pa /cm199798/cmhansrd/vo980506/debtext/80506-05.htm

House of Commons. Duty Free Shopping, Research Paper 99/74, Westminster, England (1999). http://www.parliament.uk/commons/lib/research/rp99/rp99-074.pdf

Miller, Delbert C. and Form, William H. *Industrial Sociology*, Harper, New York, N.Y. (1951)

Mayo, E. *The Human Problems of an Industrial Civilization*, MacMillan, London, England. (1933)

Martland, Carl D.; Lu, Lexcie; Shi, Dalong and Sussman, Joseph M. Performance Based Technology Scanning for Intercity Rail Passenger Systems. Presented to TRB 82nd Annual Meeting, Washington, D.C. (2002)

Marique, Justo and Jensen, Helen H. Working Women and Expenditures on Food Away-from-Home and At-Home in Spain. *Journal of Agriculture Economics*, v.49, no.3, p.321-333 (1998).

Mills, Edwin S. What Makes Metropolitan Areas Grow?

Normandau Associates. Willingness-To-Pay and Expenditures by Anglers in the Snake River Basin in Central Idaho, Drumore, Pennsylvania (1999).
http://www.nww.usace.army.mil/lsr/reports/sportfish/central_idaho/wtp_anglers/anglers.htm

National Statistics Online. United Kingdom Time Use Survey, Edinburgh, Scotland (2000).
http://datalib.ed.ac.uk/EUDL/UKtimeuse.html
http://www.statistics.gov.uk/timeuse/key_findings.asp
http://www.statistics.gov.uk/timeuse/default.asp

Spitz and Adler. Presentation to TRB 82nd Annual Meeting, Washington D.C. (2002)

Stanford Working Paper – http://faculty-gsb.stanford.edu/wolfers2/POLECON230 /documents/Session_06/Aldeasa(B)%20-%20Integrated%20Strategy.pdf

Santucci, Joseph. The Hot Times on the High Iron, Kankakee, Ill. (2001).

Small, Kenneth A. and Winston, Clifford. Ch.2: The Demand for Transportation: Models and Applications, in *Essays in Transportation Economics and Policy*, Brookings Institution, Washington, D.C. (1999)

Sousa-Poza, Alfonso; Schmid, Hans and Widmer; Rolf. The Allocation and Value of Time Assigned to Housework and Child-Care: An Analysis for Switzerland. *Journal of Population Economics*, v.14, no.4, p.599-618 (2001).

John O. Ward & Associates. Dollar Value of a Day, ExpectancyData.com, Prairie Village, Kan. (1999). http://www.expectancydata.com/dvd.html

Chapter 4

High Speed Rail Technology and Route Planning Fundamentals

This chapter reviews the current state-of-practice in high speed rail technology and route planning in the United States. In particular, attention is devoted to the following three areas: (1) How the high-speed rail routes in terminal areas are determined, using currently proposed schemes in North America and existing arrangements in Japan and Europe as case studies; (2) What is the role of magnetically levitated ground transportation (Maglev) in intercity passenger and freight transportation, and how it might fit into the current strategic planning for high speed ground transportation (HSGT), using the deployment of the Tokaido Shinkensen in Japan in the 1960s, and the construction of Shinkensen-type high-speed rail in the Formosa in the 2000s, as case studies in quantum improvement of radical technologies.

With respect to route planning, we emphasize the importance of access time and control of access mechanisms to the competitiveness of high speed rail. With the Performance-Based Technology Scanning (PBTS) framework introduced in Chapter 2, we demonstrate that very high speeds of 150mph~300mph available with expensive exclusive right-of-way technologies can achieve very high market shares at very high costs, but are not necessary and not cost effective ways of accomplishing mass intercity transportation. We postulate that Maglev technologies are most effective when deployed in an incremental fashion, and current high-speed rail strategic planning is too fragmented. Airport access, urban transit, regional rail systems, conventional high speed rail and Maglev-type technologies are really all different facets of intercity mass transportation. These different modes should be evaluated and designed together such that system optimal is achieved. In some cases, this may involve joint-operation and sharing of large capital facilities, such as track and stations.

4.1 General Review of Current High Speed Rail Planning Studies

In this section, we will review very briefly the current high-speed rail strategic plan in North America, and compare it to the high-speed rail vision of Japan in the 1960s, France in the 1970s, Britain in the 1980s, and Germany in the 1990s. Review of specific high-speed rail schemes such as the California, Midwest, Florida, North Carolina, Baltimore & Washington Maglev, Pittsburgh Maglev schemes, will be deferred to a later chapter.

4.1.1 Federal Railroad Administration Next Generation High Speed Rail Technology Demonstration Program

The Federal Railroad Administration (FRA)'s Next Generation High Speed Rail Technology Demonstration Program (NGHSR) today appears to be focused on the development of enabling technologies. The Five-Year Strategic Plan for Railroad Research, Development and Demonstrations (2002) devotes Chapter Five to the programme, which is divided into four sections: Positive Train Control, High-Speed Non-Electric Locomotive, High-Speed Grade-Crossing Protection, High Speed Track and Structures. This is not surprising, since the FRA's statutory mandate is to ensure safe operation of railroads, but not necessarily to promote the railroad industry.

The Positive Train Control programme is mainly intended to deliver a more cost-effective cab signal technology than the current state-of-practice on the Northeast Corridor, which is a system based on coded track circuits first developed by the Pennsylvania Railroad. Demonstrations are being conducted using digital radio communication technologies in the Michigan Corridor for operating speeds of up to 110mph. The High-Speed Non-Electric Locomotive Technology programme is mainly an attempt to develop high-speed locomotives using alternative technologies, including gas-turbine and flywheel, to give rise to a lower-weight locomotive with better acceleration. No attempts are currently being made to build advanced diesel locomotives (such as the British Intercity 125) to achieve the same functional performance. The High-Speed Grade-Crossing Protection programme is mainly concerned with physically preventing impatient motorist from jumping the crossbucks as an alternative to full-grade separation. The main technologies being considered are four-quadrant gates, and full-width drag-nets. The High-Speed Track and Structures programme aims to use innovative methods to reduce the cost of high-speed, freight-compatible track structures, both in terms of construction and maintenance. Attempts are also being made to address ride-quality concerns.

4.1.2 Intermodal Surface Transportation Efficiency Act Section 1010 Designated Corridors

The Five-Year Strategic Plan builds upon prior work which had already been carried out on high-speed corridor planning as part of the Intermodal Surface Transportation Efficiency Act (ISTEA). Section 1010 of the act identified five corridors, which were supplemented by three additional corridors in Section 1103(c) of the Transportation Equity Act for the 21st Century (TEA-21). These are: the Pacific

Northwest Corridor, linking Eugene, Oregon and Vancouver, British Columbia through Seattle Washington; California, linking Los Angeles, San Diego, Sacramento and the Bay Area; the Chicago Hub linking St. Louis, Missouri, Minneapolis, Minnesota, Milwaukee, Wisconsin and Detroit, Michigan; the southeast, extending the northeast corridor through Charlotte, North Carolina, Spartanburg and Greenville, South Carolina, to Atlanta and Macon, Georgia and linking Raleigh, North Carolina to Columbia, South Carolina, Savannah, Georgia and Jacksonville, Florida; and Florida linking Miami to Tampa via Orlando. The new 1103(c) corridors are: the Gulf Coast corridor linking Houston, New Orleans and Mobile plus New Orleans to Meridian and Birmingham; the Empire Corridor linking New York to Buffalo via Albany; and the Keystone corridor linking Harrisburg and Philadelphia, Pennsylvania.

Sporadic references to the ISTEA corridors can be found in a variety of documents that discuss high-speed rail planning, especially in literature distributed by high-speed rail advocacy groups. As evident from this quote, achieving the Federal designation status is a symbol of having "arrived" in the North American high-speed rail scene:

> Pennsylvania gained a designated HSGT corridor with the enactment of the Transportation Equity Act for the 21st Century (TEA-21). The newly-designated Keystone corridor, which runs between Philadelphia and Harrisburg, is owned by Amtrak, is fully electrified, and contains very few grade crossings. With official designation, dedicated Federal authorization for a Philadelphia-Pittsburgh HSGT study, and the current efforts by the State, there are now expanded opportunities to improve the speed and service of east-west passenger rail in Pennsylvania, to complement the superb north-south Northeast Corridor service already in place and constantly being improved.
>
> -- *Pennsylvania's Role in High Speed Rail, Federal Railroad Administration*

The designated corridors seem to reflect the results of political discussion as apposed to demand-based transportation planning. In the original process for selecting Amtrak routes, the Secretary of Transportation simply designated endpoints of transcontinentals, which left many important intermediate cities without direct service. The corridors designation followed a similar process which was based on historical rail traffic levels, instead of a detailed analysis of current highway travel patterns.

None of the four studies reviewed (Southeast Corridor study, Midwest Corridor study, Boston to Montreal study, Florida study) demonstrated a demand-driven way to determine routing, or showed multiple routing options. Studies that are at an advanced stage and involve selecting a route amongst different options tend to take the form of selecting a route given that a fixed list of station locations. No attempts were demonstrated to study demand variation within metropolitan areas, or refining designated routes and corridors on the basis of expected travel patterns. The current DOT high speed rail strategy seems mainly driven by a political process at a strategic level, and not a city-to-city search for the route with the highest benefits.

There also appear to be a general lack of planning resources at the local level. In FRA's NGHSR enacted budget (2002), corridor planning represents 21% ($5.9 million) of the total budget; in the 2003 budget request, corridor planning receives a mere 7% ($1.7 million). Track structures -- a major part of costs of high-speed rail systems, received an average of 5% of total budget over the two years (2002-03).

The maps that are produced by the planning process shows a number of nodes, but no intermediate points, and treat metropolitan areas as single nodes. For instance, Chicago is one node, Boston is one node, New York is one node. There is also little coordination between the different corridors: The Empire Corridor to Buffalo and the Keystone Corridor to Pittsburgh are both designated corridors; both of the ex-New York Central alignments (via Detroit and via Toledo) are designated corridors. Apparently it was deemed unwise to connect Cleveland to Pittsburgh or to serve Detroit via Toledo (or Toledo via Detroit!). Some other designations that appear suboptimal include the designated New Orleans-Houston v.s. the undesignated Texas Triangle, Houston-Dallas Ft. Worth and Houston-San Antonio. The track mileage required to serve New Orleans-Houston is approximately the same as that from Dallas Ft. Worth-San Antonio via Houston.

4.2 Detailed Analysis of An Example High Speed Rail Plan

The Boston-to-Montreal High Speed Rail study is in a very preliminary stage. We chose the Boston-and-Montreal study as an example due to the availability of data, and the cross-border international significance of the line. Other high speed rail studies in the United States may be subject to some of the same issues. This analysis is intended to serve as an example of some of the wider issues that should be considered when considering implementation of high speed rail corridors.

The Boston to Montreal Corridor (B&M HSR) is 325 miles long (309 miles by highway, MapQuest.com). Without considering whether such speeds are in fact attainable, the journey time at an average speed of 90mph is thus 3 hours and 36 minutes, compared to driving times of around four and a half hours. The principle cities visited by the B&M HSR are: Boston (Mass.), Nashua, Manchester, Concord (N. Hamps.), White River Jct., Montpelier, Burlington (Vt.), and Montreal (Que.). There is, in fact, a current Amtrak service that covers parts of the B&M HSR route from White River Jct. to St. Albans that attained an annual ridership of 76,784 passengers (NARP, 2001), or 210 passengers daily.

The study website compared the B&M HSR to the Northeast Corridor between Boston and Philadelphia. Let's see if there is anything wrong with that picture. Firstly, a demand model of the Amtrak Northeast Corridor, using very simple gravity models and Census data on metropolitan area population, is constructed. The spreadsheet analysis is shown in Plate 4.1:

Gravity Model Constant			2.50E+07								
Train Average Speed			90								
Seats per Train			200								
					Annual Ridership						
		Distance	Journey Ti	Population	BOS	PVD	NHV	STM	NYP	EWR	Sum
Boston Me	BOS	(miles)	(hours)	3297201							
Providence	PVD	50	0.56	1125639	267225						267,225
New Haver	NHV	102	1.13	522279	40785.6	20749.33					61,535
Stamford	STM	41	0.46	332835	20470.08	9431.8	15263.36				45,165
New York	NYP	40	0.44	8712600	443853.6	192929.4	202240.4	260987.2			1,100,011
Newark Me	EWR	10	0.11	1954671	95480.64	41041.01	40386.6	46841.97	PATH		223,750
Philadelph	PHL	89	0.99	4949867	176971.5	71128.89	51704.23	42668.78	1568226	391362.9	2,302,062
Total				20895092							3,999,748
								Riders per day (both directions)			10958.21
								Trains per day (in each direction)			27.39554

Plate 4.1: Simple Model calibrated with Boston-Philadelphia Amtrak Ridership Data

We calibrate the model to give us roughly the correct number of annual riders and trains per day, using important variables such as average train speed, seats per train, and a constant. This model gives us a reasonable number of four million annual riders between Boston, New York and Philadelphia. The actual current Northeast Corridor ridership is about 6.2 million for all services between Boston, New York, Philadelphia and Washington (NARP, 2001), excluding the Clockers. We don't expect this model to be sensitive to important attributes like service quality, on time performance, and locations of stations. However, we're only using this model to investigate how much demand we think there is in the corridor. It is simply intended as a screening tool to investigate the viability of corridor concepts. As a sanity check, this model was validated against the current Amtrak Vermonter ridership-performance statistics:

Gravity Model Constant			2.50E+07								
Train Average Speed			40								
Seats per Train			200								
					Annual Ridership						
		Distance	Journey Ti	Population	NYP	STM	NHV	SPG	LEB	MPR	Sum
New York	NYP	(miles)	(hours)	8712600							
Stamford	STM	35	0.88	332835	NEDirect						0
New Haver	NHV	36	0.90	522279	NEDirect	NEDirect					0
Springfield	SPG	58	1.45	573940	62021.83	3251.529	8269.153				73,543
White Rive	LEB	133	3.33	12395	659.4973	29.07834	54.22951	85.58179			828
Montpelier	MPR	59	1.48	7686	333.7821	14.31144	25.69111	36.76086	2.583538		413
Burlington	BTV	39	0.98	165917	6424.749	271.867	479.7507	659.5768	33.57618	52.31746	7,922
Total				10327652							82,706
										Riders per day (both directions)	226.5914
										Trains per day (in each direction)	0.566478

Plate 4.2: Simple Model validated against New York-Burlington Amtrak Ridership Data

With an average speed of 40mph, having excluded the riders between New York and New Haven (which are attributed to the Northeast Direct service group), the model gives a number close to the actual performance: ridership audit gives the actual number as 76,784 (NARP, 2001), v.s. projected number of about 83,000. These analyses, surprisingly simple, but give us reasonably accurate results. Considering that the model was not calibrated in any scientifically defensible way, but was based on engineering estimates of constants coupled with simple linear functions, the model appears to be robust even when applied to a corridor of a very different characteristic. So, applying the model to the Boston-Montreal corridor:

Gravity Model Constant			2.50E+07								
Train Average Speed			90								
Seats per Train			200								
					Annual Ridership						
		Distance	Journey Ti	Population	BOS	MHT	PSM	LEB	MPR	BTV	Sum
Boston Me	BOS	(miles)	(hours)	3297201							
Mancheste	MHT	53	0.59	190332	42626.91						42,627
Concord	PSM	17	0.19	38981	6610.021	1571.153					8,181
White Rive	LEB	60	0.67	12395	1131.752	110.2986	28.99017				1,271
Montpelier	MPR	59	0.66	7686	482.7102	38.72361	9.06377	5.812961			536
Burlington	BTV	39	0.43	165917	8637.816	649.6316	147.3633	75.54641	86.61994		9,597
Montreal	YMX	97	1.08	1016000	37107.21	2559.406	559.1251	232.4921	206.7082	6256.268	46,921
Total				4728512							109,134
										Riders per day (both directions)	298.9962
										Trains per day (in each direction)	0.747491

Plate 4.3: Simple Model applied to Boston-Montreal Market

The model tells us that annual ridership of about 109,000 on the corridor can be expected, or about 300 passengers per day in both directions. Since one train will carry 200 passengers, 0.75 trains per day, or one train every 32 hours could carry this load.

Is this analysis credible? The Boston & Montreal is a Federally-designated high speed corridor, and the present analysis is claiming that it will only carry some 150 passengers a day in each direction – fewer than current average ridership between New York and Chicago on Amtrak's Lake Shore Limited (based on a 18-hour one-way, non-high-speed service). Surely a Federally-designated corridor should perform better?

In many markets, the current traffic carried by existing air service is often a good indicator of how successful high-speed rail service could be. Air service could be fast and frequent, but is marred by long access times in a mid-distance market. Rail service usually attract between about 20% and 500% of riders compared to air service – for instance, between Boston and New York, high-speed rail carries some twice as many riders as the two air shuttles combined, while between Chicago and Milwaukee, air carries about three times as many local passengers daily as air service. The present model predicts that in the Boston-Montreal local market, about 37,000 passengers annually would be carried by the proposed high speed rail, or about 100 passengers per day. Examining the current airline operating plan in that same corridor, Air Canada and Delta Airlines currently operate eight regional jet flights daily, each carrying 40 passengers between Boston and Montreal. Assuming a load-factor of 70%, and a typical ratio of originating v.s. connecting passengers for hub airlines of 50%, these flights carry a total of 112 Boston-Montreal passengers per day. The Boston & Montreal will carry about as many passengers per day as air service, according to the present model. Even if the model is based on very simple linear functions, it predicted a number which is consistent with high-speed rail experience everywhere. Any rational analyst would have serious reservations believing that the Boston & Montreal would generate anything remotely approaching Northeast Corridor volumes (some 1,000~1,500 passengers daily between Boston and New York alone). By the same token, any serious analyst would have problems believing that Boston & Montreal deserves similar levels of Federal funding as the Northeast Corridor, even after the economic redevelopment potential is accounted for.

The current model does not take into account of induced demand, the economic development effects that will come about as a result of the high speed rail, or even intermodal trip opportunities that will be created by people driving from elsewhere in New Hampshire and Vermont to the high speed rail line and taking the train into Boston. Even if we assumed that these effects combined will give us four times the demand than the model projected, the load would still only justify three trains daily – probably a morning train, an afternoon train, and an overnighter extending through to Washington or elsewhere in Canada. It is unlikely that the infrastructure costs incurred in upgrading the rail line could be justified by a grand total of six passenger trains in either direction daily.

Boston-Montreal is a low-density corridor, unlike the Northeast Corridor, which is a high-density corridor. The cities between New York and Washington may have owed their development to the Pennsylvania Railroad and the Corridor, but that was a long time ago. Building a high-speed rail through Vermont and New Hampshire will not achieve the same effect, because there is now a formidable competitor providing transportation services in the region: the Interstate 89. There are probably better places to spend high-speed rail money than in the Boston & Montreal corridor. It's hardly even a corridor – nobody lives there.

There are subtle points that this model does not capture, but this model was not designed to be a comprehensive evaluation of the possibilities of the corridor. It was intended to demonstrate the type of problems associated with a strategy that focuses on a political planning process. The Boston-Montreal High Speed Rail Study Group, perhaps due to the Federal designation, does not address some fairly basic issues in project evaluation. For instance, does it make sense to simply extend the Northeast Corridor through Boston and terminating at Manchester, New Hampshire? If the Northeast Corridor were to be extended, is Manchester, New Hampshire the best destination? What about Portland, Maine? If the Northeast Corridor were to be extended, how should it traverse downtown Boston? It can go via the Grand Junction Railroad, by-passing the downtown, or the North-South Rail Link could be built, connecting the two mainline stations in Boston. Should the North-South Rail Link go through the airport? North-South Rail Link is an expensive scheme – does it make more sense to expend those funds in Connecticut and accelerate the upgrade of former New Haven trackage? These are important system design questions, which are omitted when certain corridors are given the designated status, and institutions are created to represent the interests of specific corridors.

To answer these questions, a high-speed rail vision, or strategy, is required. The vision would define the critical issues in high speed rail planning, and would lay out what the high-speed rail service seeks to provide. High-speed rail service is not good at providing service to spread out areas of small demand density; it is better at connecting points of concentrated demand density. The strategy would recognize this and provide ways to make grand schemes like the Boston & Montreal more cost-effective. High speed rail is much more than drawing straight lines between cities – it's about getting people from where they are to where they want to go. High-speed rail planners should recognize that they are not constrained by existing corridors and that they need to examine options beyond taking an already defined corridor and determining the level of investment required to generate the ridership target they have in mind.

What I have provided in this section is only a screening analyses using very general data and without consideration of important micro-effects. In the rest of the chapter, we will explore some of the questions raised in the previous paragraph. But first, we review the high-speed rail strategy and policy (whether inadvertent or deliberate) of other regions of the world.

3.3 How They Really Built the Shinkensen

A common myth amongst passenger rail activists is that Shinkansen is the model high speed intercity railroad, and that if you get the speed high enough, everybody would ride the rails, the system would be successful like the Shinkansen, and the rail line would make both a profit and economic sense. This is clearly not true, not even in the case of the Shinkansen. The original Tokaido Shinkansen, now operated by Central Japan Railway Company (JR Central), was essentially a capacity relief scheme. The later Shinkansen extensions reflect political will to connect the country with Shinkansen, and resulted in poor investment decisions that eventually bankrupted the Japan National Railway.

During the building of original Tokaido Shinkansen, a very unlikely set of circumstances occurred: (a) The Tokaido Main Line, busiest line in Japan, was facing capacity problems; journey times were slow but yet the trains were filled to the brim, to the extent that four-tracking the Tokaido Line was being seriously studied as an option; (b) In postwar Japan, there was a great sense of optimism and will to realize economic growth; (c) The existing railroad infrastructure in Japan was dilapidated and run down, due both to the effect of the War and the narrow gauge lines were never designed to handle the traffic volumes, tonnages, and speeds that were being asked from it (Yamanouchi, 2000).

Thus, the Tokaido Shinkansen was justified on the following basis: given the choice between four-tracking the narrow-gauge Tokaido Main Line, versus the construction of a set of standard gauge lines alongside the Tokaido Main Line alignment specifically for express passenger service, the incremental benefit from halving the journey time (from six-hours to three-hours between Tokyo and Osaka) and the performance gain on the narrow gauge lines (avoidance in having to pass high-speed express and freight trains) was greater than the difference in costs between two extra narrow gauge tracks and the two standard gauge tracks (see Debassay, 2003 for a detailed discussion).

The important feature to note in this case is that not even a 50% reduction in end-to-end journey time was able to justify the cost of the brand-new alignment. In this case, the reduction in journey time was only able to justify the difference in cost between narrow gauge and standard gauge infrastructure, laid down alongside the existing rights-of-way in most places by widening. Obviously, the original Tokaido Shinkansen is an astounding success – but only because the traffic was already mature and overflowing, and essentially no new rights-of-way were required. The original Tokaido Shinkansen was built in the part of Japan with the least difficult terrain, and did not enter urban areas through elaborate viaducts and tunnels; it entered urban areas by widening and realigning existing narrow-gauge infrastructure to utilize space more effectively in the existing rights-of-way.

There are many high-speed rail schemes currently in progress in Asia and Europe. Some will succeed, and some may not fare as well as the promoters hope. The on-going scheme in Taiwan bears all the hallmarks of the original Tokaido Shinkansen: the existing 70mph narrow-gauge lines are capacity constrained, especially by local commuter trains; the highways are congested, and the distances are too short and the weather too unpredictable for domestic aviation to be a major player; the new infrastructure will be constructed through mostly coastal farmland, especially in the southern part of the island; the new infrastructure will half the journey time from five hours to about two and a half hours.

On the other hand, the Channel Tunnel Rail Link, currently being constructed in Southern England, seeks to upgrade a capacity-constrained 100mph line to 183mph standards by constructing a brand-new alignment through heavily populated London suburbs. The journey time savings associated with the new alignment is in the order of 60 minutes, cutting the London-Paris journey from three and a half hours to two and a half. Apparently a good case for investment, it differs from the original Tokaido Shinkansen in that people are not exactly flocking to travel from Paris to London, or vice versa. The suburban rail congestion could be relieved with advanced signalling technologies, without resorting to a brand new alignment. A study by Virgin Rail suggested that Regional Eurostars (from Glasgow, Manchester, Birmingham and other cities to Paris) would only be justified with a frequency of one train per day from Glasgow, and two trains per day from Manchester. The Channel Tunnel Rail Link is likely to see a maximum of ten trainloads (perhaps a maximum of about 8,000 passengers) in each direction per day, compared to the inaugural hourly service (18 trains per day) planned for the Taiwan high speed rail, and the seven-minute rush-hour headways that currently run on the Tokaido Shinkansen. The economics of the Channel Tunnel Rail Link pales in comparison.

3.4 Conclusions

In this chapter, we have illustrated a few fundamental principles of high-speed rail planning through case studies, logic, and literature review. First and foremost is the important realization that high-speed rail planning is a detailed exercise. Cities cannot simply be treated as nodes of different sizes, since at the distances where high-speed rail is most likely to be competitive (i.e. under 300 miles), the access time to and from the central station could be a significant part of the total journey time, and factors into competitiveness against the other modes very significantly. The current status quo of designating specific corridors to be studied as high-speed rail corridors could lead to undesirable results by allowing a fixed-route or fixed-endpoints focus to develop, when a rational analysis of which nodes to include, what route to take, and where to stop, is needed. Secondly, in terms of infrastructure investment, it is extremely difficult to justify new infrastructure based on journey time savings. Proposals for new infrastructure, new technologies, or new rights-of-way are much more likely to succeed where capacity is a constraint. Economic analysis showed that the age-old rule of investment decision: 'buy the best you can afford, then use it until it claps out' seems to apply to high speed rail also. The unfortunate result for high-speed rail advocates is that quantum speed improvements are unlikely to occur unless a situation exists where the existing infrastructure is already handling as many trains as it possibly can. In the case where existing infrastructure is unable to attract sufficient riders to justify its operational costs, investment in new infrastructure is likely to fail also.

A full analysis of the implication of this result on magnetic levitation technology appears in the Lu, Martland and Sussman paper (2003) – see Appendix H.

References

Armstrong, John H. The Railroad What It Is, What It Does. Simmons-Boardman Books Inc., Omaha, Nebraska (1995).

Cascetta, Ennio. Modelling the Spatial Interactions between Socio-Economic Activities and Transportation System: Two Case Studies, *Presentaion to MIT Center for Transportation Studies*, Boston, Mass. (2002).

Debessay, Wintana. A Case Study on the Shinkansen: The Effect of Political Overlook. *Proceedings of 1.011 Project Evaluation Term Project Presentations*, Massachusetts Institute of Technology, Cambridge, Mass. (2003).

Federal Railroad Administration. Five-Year Strategic Plan for Railroad Research, Development, and Demonstration, Report No. FRA/RDV-02/02, Washington, D.C. (2002).

Midwest High Speed Rail Coalition. Midwest High Speed Rail Report **9:**4, Chicago, Ill. (2002).

McCown, Robert J. Next Generation High Speed Rail Progam Status Overview, Office of Railroad Development, Federal Railroad Administration, Washington, D.C. (2002).

National Railroad Passenger Corporation. Amtrak National Timetable, Fall 2002/Winter 2003, Philadelphia, Penn. (2002).

Roth, Daniel L. Incremental High Speed Rail in the U.S.: Economic and Institutional Issues, MIT Thesis, Cambridge, Mass. (1994).

Sussman, Joseph. Chapter 1 in *Introduction to Transportation Systems*, Artech House, Boston, Mass. (2001).

Tam, Ryan. Air Service and Economic Development: The Case of American Airlines at Raleigh-Durham and Nashville, Strategic Location and the Changing Corporation, Harvard Graduate School of Design, Boston, Mass. (1998).

Transportation Research Board. Working on the Railroad: Getting Passenger Services Up to Speed, *TR News*, No. 222, Washington, D.C. (2002).

Yamanouchi, Shuichiro. If there were no Shinkansen. International Department, East Japan Railway Company, Shibuya-ku, Tokyo, Japan (2000).

Chapter 5

Case Studies in Downtown Access Design

This chapter is both a review of existing schemes for intercity rail's access to the downtown core, and an application of the Performance-Based Technology Scanning framework (See §2.10) to designing an adequate distribution system for a large metropolitan area with a population of more than one million. The central idea is that when evaluating high speed rail strategies, it is easy to get carried away with maximizing speed when the most leveraged infrastructure investment could be in the downtown, especially when high speed rail schemes are considered jointly with commuter rail and rapid transit improvements.

There are a number of urban distribution system designs for intercity transportation: (1) travellers could transfer to the local public transit system; (2) travellers could transfer to local public or private transport at a central collection point; (3) travellers could transfer to local private transport at a number of collection points. In this chapter, we use the Performance-Based Technology Scanning framework to demonstrate that alternative (3) would make intercity rail most competitive with alternative modes of highway and airline. Given the same technical capabilities and amenities, alternative (3) would result in the longest "breakeven" distance; given an origin-destination market, alternative (3) would allow lower costs by requiring lower speeds for a given target mode share. In transit-oriented cities, (1) can appear to be the most cost-effective solution, but the de-facto need for most target customers to make two or three transfers makes it unattractive.

In the final part of the chapter, we review a number of actual examples where this concept has been applied to a number of cities. While no attempt was made to survey the examples in a level of detail required for engineering, attention was devoted to existing infrastructure and demand patterns to enable evaluation of benefits and whether this approach makes sense for the cities concerned.

5.1 Review of Intercity Rail Distribution Systems

As part of research carried out on the behalf of the Union Internationale des Chemins d'Fer (UIC), the author reviewed urban rail networks in many of the world's metropolitan areas. The coverage was not intended to be exhaustive but aimed to cover a representative subset of downtown distribution designs. The detailed case studies were previously published in a working paper submitted to the sponsors (Martland et al., WP-2002-02). The following is a representative summary of the relevant sections.

5.1.1 Tokyo, Japan

The non-Shinkensen rail in downtown Tokyo are integrated into the extensive city subway network, and there is track sharing and vehicle interchange between the Tokyo Municipal Subway and some of the regional rail operators. The vehicles are interoperable on both types of the infrastructure. Like other subways and regional rail networks in very large metropolitan areas, Japan has a nearly totally-connected downtown rail system. The incoming regional/intercity train will often visit a number of stations before finally terminating at Tokyo Central. Some trains run through the city. Real estate experience over the most recent decade suggests that other business districts have developed around rail stations that are not in the downtown, and a substantial number of regional rail riders travel directly to their target business district without transferring at the central station. Connecting eight of the most important business districts within the Greater Tokyo region is a circular rail line called the Yamasaki line. Shinkensen Expresses continue to call only at the central station.

5.1.2 London, England

Downtown distribution in London is something of a historical accident. In the era of private railways, each long-haul railway constructed its own magnificent London terminal and transferring between lines often required crosstown connections on the Underground network. In the downtown, tracks were not generally shared between full-gauged mainline trains, small-gauged surface line trains, and tiny-gauged deep level tube trains. Reaching the final destination after reaching the London terminal of the line in question was often difficult, involving multiple transfers or a congested hackney carriage ride. Beginning in the late 1980s, British Rail recognized the need of crosstown travellers and instituted the Thameslink scheme which called for construction of a new elevated railway across London to provide a North-South connexion. Additional stations were also provided. Today, two other schemes, named Thameslink 2000 and Crossrail, are in progress to open more direct cross-town links for mainline trains.

5.1.3 Harbin, Heilongjiang, China

The downtown distribution facilities in Harbin have historically been based on the London model, where tracks from the South and from the North each terminated at separate stub-end termini. There are four rail stations within the urban area, linked by a circular railroad, but two (Binjiang and Xiangfang) are mainly loose-car freight stations, while the two passenger stations deal with diesel (Northbound) and electric (Southbound) trains respectively. The 2002 operating plan and track speeds do not allow the circular railroad to act as a collector for long-distance trains in a competitive manner; passenger trains traverse round the ring mainly due to operational reasons. However, in Harbin, this is less of a problem than one might imagine, due to the tendency towards highly compact planned cities in former communist China. There are a number of business districts in Harbin, but the main station is the only station close to the largest and busiest part of the downtown; the freight stations lie in industrial areas and other stations lay beyond the main conurbations. Thus, the fact that Harbin is still a single-node city isn't a major problem, since the population is also highly concentrated, unlike in the developed world.

5.1.4 Philadelphia, Pennsylvania

Philadelphia's regional rail system has three main stations in the downtown: 30th Street, Suburban, and Market East. Prior to 1984, the Reading Railroad and the Pennsylvania Railroad independently operated Broad St. and Suburban regional rail stations which were not interconnected, while 30th Street was exclusively reserved for long distance trains. The downtown tunnel, constructed for $300 million, permitted abandonment of Broad St. and allowed access to all regional rail lines from all three stations, including a brand-new interchange with the transit system at Market East. An indoor mall connected 8th & Market transit stop with Market East regional rail station. Access time to regional rail service from the downtown was greatly reduced with the tunnel. Regrettably, due to existing infrastructure and other constraints, intercity trains continue to call only at 30th Street, including commuter trains from New Jersey. Although reaching downtown is an easy transfer, access from other significant demand generators in the city neighbourhoods is problematic.

5.1.5 Boston, Massachusetts

Boston's intercity rail system has evolved much over its 150 years' history. As the cultural center of New England, initially the London model was adopted. Major restructuring during the 1890s by the Boston & Albany and the New York, New Haven and Hartford saw the consolidation of many smaller stations into two stations, Back Bay and South Station, which both served all of the south side lines. Unfortunately, the Boston and Maine kept its own North Station for the north side lines. In the Amtrak era, a suburban park-and-ride station known as Route 128 was added. Later, two stops within the city, Forest Hills and Ruggles, were added to the regional rail network to serve major demand generators besides the downtown and provide transit connexions. In general, accessibility of regional rail service is much better in Boston than in many other American cities. Currently, there are proposals to link the North and South stations by a downtown tunnel to create a Philadelphia-like layout, and new regional rail stations that provide transfer opportunities away from the downtown are in the process of being created as part of the Urban Ring project.

5.2 Some Basic Concepts

5.2.1 Why is Urban Transportation Relevant to Intercity Rail Providers?

After the public takeover of urban mass-transit systems occurred on a wholesale scale during the 1950s, there was a growing body of opinion amongst intercity carriers that local distribution was the responsibility of the local government, and while intercity carriers' responsibility terminated at the local access point for that particular mode of transportation, whether than be an out-of-town airport, a downtown railhead, or a remote freeway interchange. Although the intercity carrier will work in conjunction with the local government to provide interchange facilities, the public intercity carrier saw little reason to enter a market in which a public monopoly was unable to operate profitably. However, there are often tremendous economies of density in urban transportation, to the extent that certain busy urban corridors may be profitable to operate independently of the local transit authority – a phenomenon evidenced by the increasing independent private investment in downtown-to-airport type shuttle buses. In this section, we argue that the intercity carriers are not only able to profit from this type of urban distribution market, but will also enhance their mode-share in the intercity market by providing better access to its services. These access issues are particularly important in congested cities where demands for intercity services are high and alternatives to the rail mode are also easily accessible.

In other words, access to the rail mode only needs to be as good as access to the competing modes such as limited-access highways or airways.

5.2.2 Local Feeder Routes in a Metropolitan Urban Setting

In cities with a large population, many modes compete with intercity rail for travel over distances of between 100 and 1,200 miles. Typically, at the "origin" end, a mode-choice tree in a major metropolis such as New York City might look something like this:

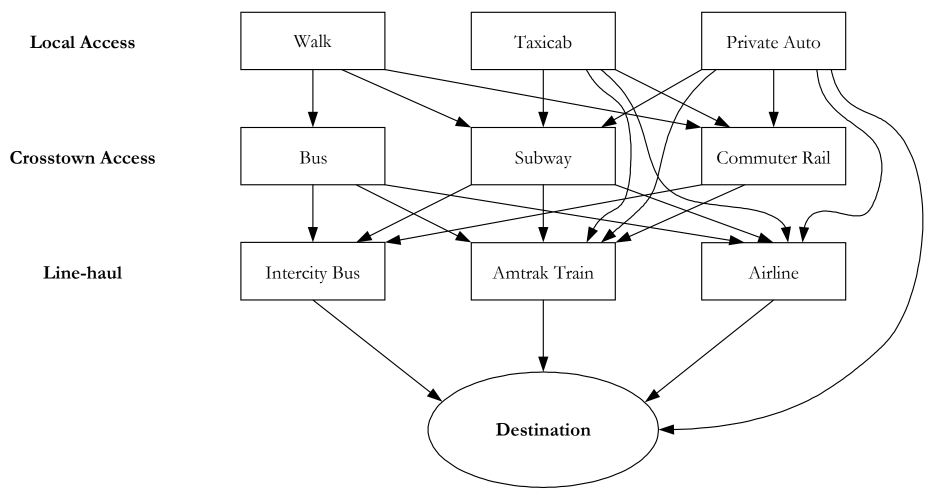


Plate 5.1: Local Feeder Routes in a Metropolitan Setting

More importantly, each of the paths in the mode-choice tree has a utility and idiosyncratic preference associated with them. Thus, when the traveller is making the decision with respect to the line-haul mode, the utility associated with the "local" and "crosstown" access modes will affect their decision. While intercity service providers seldom control local access, they should be aware of its importance.

In order to encourage patrons to choose their mode on the line-haul portion, a carrier must consider not just the line-haul leg, but also the access legs, so that overall utility for the journey experience for the line haul plus some combination of the "access" modes comes out to be the highest. In major metropolises, airports are frequently easily accessible using the local transportation infrastructure, as are intercity bus terminals. For that reason, the rail operator needs to choose a routing and stopping pattern carefully to enable the best downtown access by patrons who may "prefer to access the line-haul mode by private auto" or "are only able to access the line-haul mode by subway".

5.2.3 Local Feeder Routes in a Rural Setting

Interestingly, this can be contrasted with a small town or rural area, where the accessibility to much of the infrastructure that city-dwellers take for granted are simply not present. In these areas (such as North Carolina in the United States), the mode-choice tree will look more like this:

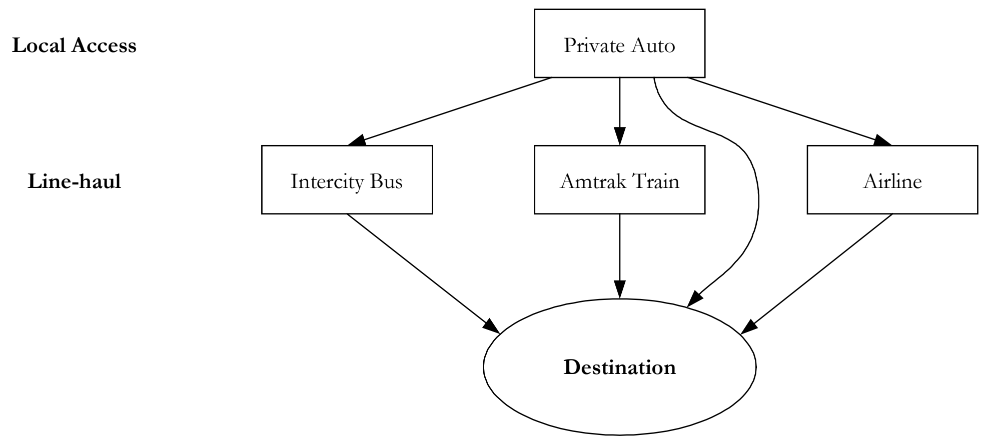


Plate 5.2: Local Feeder Routes in a Rural Setting

The key observation here is that in working to improve access to intercity rail travel, the access only needs to be as good enough to make the overall trip experience better than that offered by the competing modes. In general, there is much more competition in the large cities both between access modes and line-haul modes, thus it is a much more difficult market to break into. In rural areas, the only access mode is the private auto; thus intercity rail can attain a larger potential customer base simply by ensuring that its stations are strategically spaced such that a drive from most points in the rural state is less than the drive to the nearest major airport in the area, and by ensuring that the disutility of transfer is more than compensated by the utility in travelling by train (versus driving directly to the destination).

In many ways the rural market is a much simpler market than the urban market. In a developed economy such as the United States, where it can be safely assumed that the majority of the rural population have access to a car, access barriers such as "an hours' drive to the nearest rail station" isn't really an issue. Rural population and suburban population in general accept that a longer drive and increased disutility of access to "urban" facilities is part of the cost of living in exclusive suburbs or rural locations. A possible reason behind the failure of railroads in developed countries to attract a significant rural ridership may lie with their inability to distinguish its product from driving directly to the destination.

5.2.4 The Coupling Between Local and Long-Haul Transportation Systems

The success of intercity transportation carriers are inseparably linked to the effectiveness of the local distribution system from the long-haul nodes. Consider Slide 15 in Appendix B. The airline community discusses airport access in many publications (for a review, see Chapter 2), and the importance of such local runs are extensively documented in literature dealing with intermodal freight. It appears that the success of intercity rail may depend on the availability and quality of the local transit system and highway access network. As noted in Appendix B, in a number of case studies and example itineraries that were examined, provided that an intercity rail line of a medium-speed standard (about 90~110mph) exists, the most leveraged investment may be to improve revenue through reducing access time by offering better access within the metropolitan areas.

5.2.5 The Folly of the Union Station

Many transit (and intercity passenger transportation) professionals have come to believe in recent years that a consolidating approach to intercity services is a good thing – many cite the Union Station's downtown location (a intercity rail "hub") as a great attraction for travelling by train rather than by private auto or aeroplane. In fact, the Union Station's downtown location is an impediment, not an attraction, to suburban dwellers who wish to depart from their home and those who work in suburban business districts. In contrast, the out-of-town airport offers much better access if the origin is within the suburbs. The originating demands for intercity travel from the suburbs and the non-downtown city neighbourhoods, when integrated across the entire metropolitan area, dwarfs the originating demands which are within easy walking distance of the downtown Union Station. Some graphical illustrations of the demand studies are shown in Slide 12 of Appendix B.

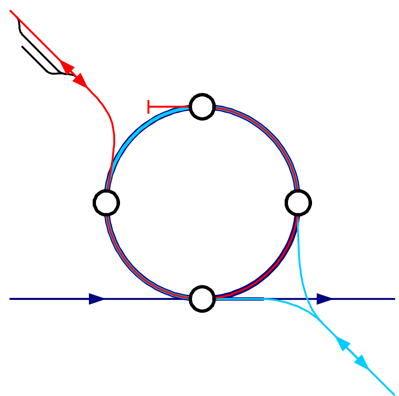
Although a beltway "Park and Ride" similar to Route 128 station in Massachusetts and New Carrollton station in Maryland solves the problem for suburban dwellers who live on or near the beltway, it does not make the intercity train any more attractive than the plane which is also similarly situated at a beltway location. In addition, the "Park and Ride" does not make it any easier for those who work in suburban business districts or city neighbourhoods that are not within easy reach of the downtown. Especially in a major metropolis, getting downtown by either transit or private auto can be a major hassle, comparable to getting to the airport. In addition, other issues exist in congested downtown that make the out-of-town airport even more competitive; for example, the lack of direct access via high-performance urban expressways to the downtown rail terminal, and the lack of affordable parking at the rail terminal, all impair the ability of the downtown terminal to serve the suburbs and some city neighbourhoods effectively.

The simple analogy with the interstate network will demonstrate why more than one downtown station is required. An intercity railroad terminating only at the Union Station and beltway Park & Rides is akin to an urban interstate expressway which has only three exits – one at the eastern intersection with the beltway, one at the downtown, and another at the western intersection with the beltway (see top of Slide 5.3), requiring the traveller to proceed through the city on slow arterial streets (akin to feeder transit systems or feeder buses to the Union Station and beltway Park & Rides). Obviously, the nature of the railroad technology requires the minimization of stops for through passengers to minimize the dwell times. However, where a train is terminating in an end-of-the-line type city (such as Boston, Mass), or a service is obviously designed for short-haul passengers (such as the Amtrak Clocker and Keystone services), increasing the accessibility of such services through a downtown distributor can make the service much more attractive than a point-to-point, airplane-like service (see middle of Slide 13 in Appendix B). It is also important to note that premium express services, such as the Acela Express, should probably not spend time picking up passengers around the loop in Baltimore, Md. or Philadelphia, Penn.; but it should be sent around (hypothetical) distributor loops in major destinations such as Boston, Mass.; New York, N.Y.; and Washington, D.C.

The goal of such downtown distributor is to bring the intercity train to within about 10 minutes' walk or taxicab ride of most parts of the city, including suburban business districts and the downtown area. Thus, the scale is important. Access time of less than 10 minutes makes a 45-minute taxicab-ride to the airport plus an hours' waiting in line for check-in much less attractive; had we retained the downtown Union Station, the congestion in the downtown area would probably make airport and Union Station access time similar in a taxicab from most suburban locations and city neighbourhoods.

5.2.6 Advantages of the Downtown Loop

The downtown loop also allows the large mega-city to be converted from a stub-end stop to a through-stop which has certain operational advantages. Decreased platform occupancy times allows more effective platform capacity utilization in an area where additional platforms may be an expensive proposition, in addition to allowing large preparation yards to be moved out-of-town where land (and possibly labor too) is cheaper. Conceivably the through services, shown in deep blue, will simply travel once around the loop, adding about 20 minutes to the through journey time (which is a lot better than the one-hour usually allowed for cross-London transfers), whilst terminating services will go around the loop before reversing at a siding or simply continue via a wye back towards the origin having completed loading and unloading at a number of intermediate stops. In the latter case, all train preparation will occur at the other end of the line, or in a yard situated in a rural area between two mega-cities. This is, in a way, the merry-go-round concept in coal transportation as applied to passenger service.



Decentralized
Terminals Design
(Service Pattern)

Plate 5.3: Decentralized Terminals Service Design

Aside from the operational and access benefits, the downtown ring has an additional benefit over other possible layouts to enhance access. Instead of building a number of lines that independently cross the city center, a ring can effectively channel traffic from any direction through the city without incurring the expense of connecting all lines to every other line. To be precise, if there are more than two connectors through the city (i.e. more than a single North-South link and a single East-West link), the ring can provide an alternative with less mileage and nearly the same through journey time and access time for those who live in the metropolitan area.

5.3 Results from The City of London Study

To illustrate the concept of better rail access for intercity carriers operating to major metropolises, a scheme was designed around the City of London to evaluate the potential benefits and feasibility of applying such an idea in an actual situation. The access scheme designed involved constructing a circular heavy-rail alignment which would connect most of the intercity terminals in downtown London. In this study, the following questions were addressed:

- What are the estimated journey time savings for connecting and terminating passengers?
- Is a double-track ring railroad sufficient to carry all the trains arriving during the off-peak hours?
- Are reasonable turn-around times for intercity trains attainable?

This is purely a hypothetical scheme, intended to demonstrate the concept and illustrate the scale of the ring in question. Of course, there are many constraints on London's radiating intercity lines which will prevent interlining between them, at least in the short term. However, as a long term proposition, the idea has some potential.

A draft diagram showing the proposed route of the scheme, which in fact closely mirrors London Underground's Circle Line, is shown on Slide 18 of Appendix B. The existing Circle Line alignment is unsuitable for use by intercity trains due to its restrictive loading gauge and other elements of infrastructure geography. The costs are thus based on a new-build scenario.

Many assumptions are necessary to complete the preliminary study:

- Thameslink and other cross-town schemes (e.g. LUL Met Line) do not exist.
- Transfer passengers must disembark at downtown London terminals, transfer via the Tube, wait for the train at the originating terminal for an "average" length of time.
- Average wait time is half the expected headway.
- Expected headway is the number of trains per hour divided by number of branches served by the downtown terminal. (e.g. London King's Cross serves the Cambridge Flyer (half-hourly), Peterborough Local (half-hourly), Leeds Intercity East Coast Train (one per hour) and Newcastle Intercity East Coast Train (one per hour), thus the average headway is six trains an hour divided by four destinations = every 40 mins).
- Walking times between Tube stations and Train Platform are Railtrack's official figures, to include waiting time on LUL during daylight operating hours.
- Walking between mainline stations is permitted, and walking times are used where possible.
- Transfer times include the walking time, Tube time, and expected waiting time for the mainline train.

The most important assumption is that crosstown schemes do not exist, since the crosstown links are in fact a part of an expensive solution that a ring-railroad would be attempting to avoid. Given that all lines are interconnected through the city, it makes the case for a ring railroad weaker. In many cities (such as Chicago, London and Boston), the lines from north side and south side are not in fact fully connected. In this study the base case of having none of the lines connected (except by transfer to a local transit system) is evaluated against building a downtown distributor to connect more stations.

5.3.1 Methods and Results from Running Time Analysis

To calculate a set of transfer times across London, it is necessary to calculate the amount of time it will take for a terminating intercity train (or indeed run-through intercity train) to travel around the inner ring railroad. In addition, running time analysis allows us to determine whether the amount of vehicle time spent in sending the train around the ring results in a saving over the turn-around time at a stub-end terminal. If the time spent around ring were significantly longer than the turn-time at terminals, it would be hard for intercity operators to justify tying up their productive asset for what is effectively a service enhancement that may or may not generate significant additional revenues. Also, if the time spent trundling around the ring is slower than what can be offered by local transportation options, the exercise becomes pointless since the added convenience of a one-seat ride will be unlikely to offset the additional time-cost of travel for many travellers. Thus, for the idea of a ring to be viable, we must fulfill the following market (competitive) and cost (operational) constraints:

- Journey Time faster than local transportation options (which may involve waiting and transfers, but will almost certainly take a more direct route)
- Running Time around the ring faster than turn-around time achievable at stub-end terminals
- Sufficient capacity on the ring to accommodate trains from most important destinations

To calculate the running times, distance around the ring was measured and an average speed achieved between any two station-stops was calculated based on a formula which took into account of the distance between stations. The formula was roughly calibrated to those sectional running times that are achieved by the London Underground's Circle Line, taking into account the added station stops served by the LUL trains but also the better acceleration and low-speed characteristics. Station dwell time at each terminal were assumed to be one minute – the stations would have to be designed as through-stations with either a center platform or two platforms with center express tracks and outer local tracks/sidings to minimize station dwell times (See Exhibit 5.4 below).

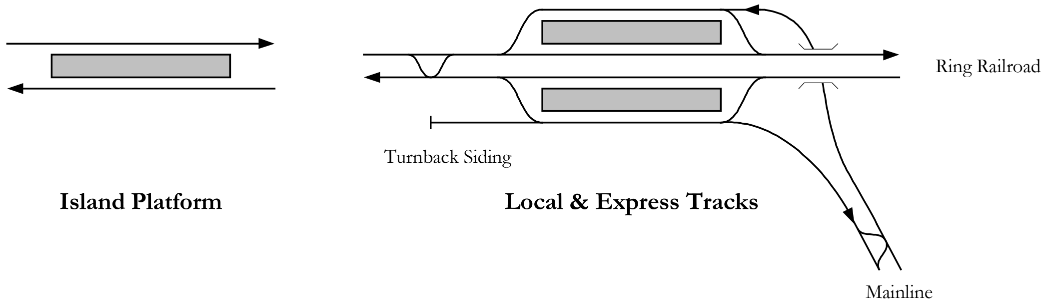


Plate 5.4: Station track layouts suitable for downtown ring railroads

The main result in this case study was that it was impossible to stop at all of the BR London terminals yet maintain a sensible running time around the ring. However, if limited-stop service was introduced, overall journey time savings are possible and the majority of passengers would still be able to make a same-platform or cross-platform transfer to a service departing in a different direction, and most parts of Central London remains directly accessible from the stations at which the intercity trains stop. In particular, the running-time for an intercity train around the loop was calculated at **51 minutes** for the all-stop scenario, and **30 minutes** for the limited-stop scenario (including schedule padding and time taken for crew changes). The current London intercity operators schedule between 40 and 50 minutes for servicing at their London terminals, although it can be accomplished in 15 minutes if an incoming train were to arrive late. The limited-stop schedule would suggest that no extra vehicle costs will be incurred by any operator if they were to send the trains around the ring; on the other hand, if the incoming arrival is more than about 20 minutes late, a decision could be made to terminate and turn the train at the operator's own London terminal in a stub-end platform.

Using the existing Train Service Database information maintained by Railtrack, and running times around a hypothetical ring, the train services' arrival times at King's Cross station was calculated and schedules created using King's Cross as a timing point. The result of this analysis demonstrates that at off-peak times, theoretically, most of the long-distance arrivals at the London terminals could be handled with a single-track ring-railroad, although obviously with a single-track loop the traffic would only be able to travel in one direction. If three-minute headways were to be maintained at King's Cross, minor adjustments of up to 6 minutes is required on some line-haul schedules, but the majority of the trains would receive a train-path on the ring. Although the practical capacity is likely to be much lower, the idea of sending most long-distance arrivals around a ring appears plausible in London provided that a double-track ring is equipped with signalling for about 20 trains per hour, roughly the limit of current technology.

5.3.2 Methods and Results from Transfer Time Model

The transfer time model is fairly straightforward. Basically, the three components of a cross-London transfer time were added together, using Railtrack recommended values. Some of the values were then changed to reflect the presence of a ring-railroad. The three components of the transfer time are:

- Walking time to and from the station platform
- Expected waiting time plus running time on the London Underground (the Tube)
- Expected waiting time for the next mainline train to your final destination

The expected transfer times from all nodes to all other nodes (including cross-platform transfers at the same node) was then calculated in a matrix. The expected transfer times were then averaged across all possible combination of nodes and this is known as the average cross-town transfer time. Although this methodology is not totally watertight (for instance, it fails to account for the fact that some transfers are heavier than others in terms of passengers, thus should be weighted more), it gives a good indication of how long one may expect to spend on an average cross-town transfer in London. Due to the difficulty of obtaining detailed transfer details, no modelling was attempted in this area. The resulting access times between every station pair are presented in Table 5.1, and summarized below:

<u>BEFORE</u>

- Average cross-town transfer time between major downtown terminals = **58 minutes**.
- Best average access time* from any one major downtown terminal = **52 minutes** (from King's Cross and St Pancras to all other stations).

<u>AFTER</u>

- Average cross-town transfer time between major downtown terminals = **47 minutes**.
- Best average access time* from any one major downtown terminal = **40 minutes** (from King's Cross and St Pancras to all other stations).

  * To all other stations, with equal weighting. The time can also be considered indicative of average transit access times from locations within the North Circular Road beltway – the walk time and wait time at the mainline station remains constant, while the Tube time and walk time from residential/business location to nearest transit stop is comparable to the walk time from stations and the Tube time.

	Cross-London Journey Time [6]	W	1	2	3	4	5	6	7	8	9	10	Station	Tube Stop
W	Walk [1]		12	12	12	15	17	15	18	15	9	10		
1	Anglia	12	25	25	25	28	59	30	37	33	48	40	Liverpool St	Liverpool St
2	ECML	12	30	30	30	30	55	32	39	35	52	56	KingsX	KGX/St Pancras
3	MML	12	40	40	40	40	65	42	49	45	62	66	St Pancras	KGX/St Pancras
4	WCML	15	28	25	25	25	50	24	31	27	58	56	Euston	Euston Square
5	M40	17	74	65	65	65	40	66	76	83	73	76	Marylebone	Baker St
6	Great Western	15	30	27	27	24	51	25	29	25	68	65	Paddington	Paddington
7	South Western/Portsmouth	18	37	34	34	31	61	29	25	26	48	60	Waterloo	Waterloo
8	Brighton	15	38	35	35	32	73	30	31	30	64	60	Victoria	Victoria
9	Kent Coast	9	56	55	55	66	66	76	56	67	33	60	London Bridge	London Bridge
10	Southend (LTS)	10	55	66	66	71	76	80	75	70	67	40	Fenchurch St	Tower Hill

Table 5.1: Cross-London Journey Times between all Mainlines, with hypothetical Ring Railroad

Because some of the minor mainline terminals are not included in the London Inner Ring Railroad scheme, the passengers arriving at those stations will continue to have to rely on the Tube for cross-London transfers. Most stations excluded were stations designed to serve only commuters. Because of this, the actual benefit realized for long-distance intercity travellers would in fact be higher than the result of 47 minutes would suggest. Although it is disputed as to whether long-range commuters or long-distance intercity passengers are the more leveraged area of the market, since this study mainly concerns intercity passengers, the assumption is made here that long-distance passengers generate more revenue per train for the railroad industry. It would not be difficult to simply substitute the set of intercity trains which we choose to send the ring with a set of long-range commuter trains, should the latter turn out to be the case.

For a conservative estimate, we shall consider the transit time from a location inside Central London to be half of the transit time between stations. We know that the access time to a Central London business/residential location is between half and one times the transfer time between London Terminals.

5.3.3 Simple Benefit Analysis

The simple benefit analysis evaluates the benefits to riders, taking only into account of the time saved for those who are already travelling by the rail mode. The actual benefits are likely to be higher as people switch to rail for its increased convenience.

Typical daily weekday flow at King's Cross = about **77,000 pax/day** (Calculated using a simple model based on average load factors on the intercity expresses and their seating capacity). Since King's Cross is an average station (e.g., Paddington, Euston and Victoria are all busier than King's Cross, while St

Pancras, Fenchurch St, and Marylebone are quieter), if we assume the typical weekday flow is an average, the typical weekday flow through London is thus approximately 930,000 pax/day.

Commuters are likely to be in high-paying jobs, thus average value of time is assumed to be £10 per hour ($15 per hour). Calculating the average value of time spent in access after an outer-suburban commuter rail or intercity rail ride, using the following passenger mix:

- 25% Cross-London Transfer pax (including Commuter Rail Transfers)
- 55% Downtown London Terminating pax
- 20% Non-CBD Terminating pax (requiring a Tube ride)

Sum saved per weekday (in terms of access time) = $1.34 million per weekday (or $350m per year). This is sufficient to support a project costing about $3.5 billion over 15 years. Some of the detailed analyses supporting these claims are shown in Slide 18 in Appendix B.

We are not trying to build a business case for the Inner London Ring Railroad. There are many external benefits which are not explicitly accounted for in this model, e.g. greater mode-share due to better accessibility, and possibility of deferring capacity enhancements on the London Underground. A closer examination of the costs and the local conditions would be needed before any firm conclusions could be drawn for London. The ring railroad is likely to more generally remain a viable option as a way to reduce access time and through journey time for any city with more than about two million population.

5.4 Application of Performance-Based Technology Scanning

How could the Performance-Based Technology Scanning (PBTS) framework be applied to designing a distribution for the core downtown? The two critical ideas in the PBTS framework are: (1) Different types of technological improvements or infrastructure investments may result in very different changes in journey experience; these changes could be evaluated with suitable assumptions regarding values-of-time; (2) Investment should be targeted in areas where the maximum return could be obtained for a given investment. Because of the complex nature of transportation systems, it is likely that the most effective investment would comprise of a package of moderate improvements rather than maximization of a single variable such as speed or accessibility.

Some emerging technologies will affect the cost-base of providing the service directly, while others will affect the revenue potential directly. Some technologies will enable service designs that were previously uneconomic to be operated -- this is a very subtle effect, which could be termed the second order impact of technologies. In this section, we design a downtown distribution system to increase revenue, which may or may not be profitable. Then we ask the question: are emerging technologies able to reduce the costs to the extent that this design becomes economic? Various answers are possible. Perhaps no new technologies are necessary, or perhaps the technological advancements required are beyond even the most optimistic current projections.

In the remainder of this chapter, a conceptual model is presented to support what was demonstrated through examples and case studies in the earlier work.

5.4.1 Assumptions of The Model

In developing the model, a number of simplifying assumptions were made. The simplifying assumptions were necessary for two reasons: (1) to keep the model tractable and implementable without specialized programming tools; (2) to retain some generality in the model, such that the results will be applicable to a wide range of metropolitan with roughly similar structures. With the necessary resources, it is possible to extend the model to cover a wider range of modes, better spatial resolution, and more detailed consideration of any or all aspects of the many variables.

There are three modes available for intercity travel: Rail, Air, and Auto. Other possible modes for an extended version of the model are: Rental Car, Intercity Bus, Air via Hub, and Auto via State Highways. There are two modes available for access to terminals: Auto and Transit. Other possible modes might be: Walk, Bicycle, Taxi, Express Bus.

5.4.2 Structure of The Model

The metropolitan area is divided into a number of quadrants of equal sizes (25, in the case of the experimental model), and the central business district is located in the middle quadrant. The terminal locations for each of the modes are then identified. In the case of air, the airport may be within a quadrant far from central city; in the case of rail, the union station may be in the quadrant containing the central business district. In an intercity rail system with a downtown distributor (MetroFlyer style, see Appendix B), three stations may be located in different quadrants of the metropolitan area. For each intercity mode, the locations of terminals are entered into a two-dimensional array (5*5, in the present set up), and the terminals are given an unique identifying number. Quadrants without terminals are assigned a value of NULL. For the auto mode, each quadrant is a terminal.

Terminal Locations (Rail)				
	1			
		2		
			3	

Terminal Locations (Air)				
				1

Terminal Locations (Auto)				
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Plate 5.5: Model Spreadsheet, Part 1

Using a simple algorithm, the nearest node to every quadrant in the metropolitan area was then determined, storing the results in a separate array. Using a shortest path algorithm based on distance, the shortest paths between each quadrant and the nearest node is calculated. The result is stored in a linked-list using an array of pointers. This array of pointers would later be used to calculate the access time.

Congestion Factors

In most metropolitan areas, significant auto congestion occurs in the downtown at most times during the period when intercity travel is most likely to take place. Even if actual congestion does not occur (i.e. when vehicle density exceeds critical density), it is likely that in the downtown area, private autos will achieve a lower average speed than in the suburban areas, due to a mixture of factors such as higher density of traffic lights, lower lane widths, speed limits, and higher probability of pedestrian interference. To accurately assess the access time, these factors need to be taken into account. If the degree of congestion (between zero and one) was assumed to be inversely proportional to the distance from the city centre, the following array of congestion factors are obtained:

Congestion Factors (0 = freeflow)				
0.25	0.25	0.25	0.25	0.25
0.25	0.5	0.5	0.5	0.25
0.25	0.5	1	0.5	0.25
0.25	0.5	0.5	0.5	0.25
0.25	0.25	0.25	0.25	0.25

Plate 5.6: Model Spreadsheet, Part 2

If the freeflow speed of private autos on arterials were assumed to be 40mph, and the quadrants were assumed to be four miles in length (giving a total metropolitan area of about 20 miles in diameter), the access times from each quadrant to the nearest node using the shortest path is calculated for each mode. Perhaps of crucial importance at this point is the need to stress the obvious fact that the access times are not directly comparable for the different modes from the same origin, since the location of the terminals may be different. The result from this part of the model is a set of arrays, one for each mode, containing the access times from each quadrant to the nearest terminal, using the private auto as an access mode.

For transits, the freeflow speed is similarly assumed to be 25mph (clearly achievable with high-quality heavy-rail transit or high-quality bus service on uncongested highways). For simplicity, a generic transit network of two core rail trunk lines radiating from the city centre in an X-shape was assumed, with bus feeders serving all other quadrants. The number of transfers required to reach the central rail station and the airport was then manually calculated, based on the assumed transit network. The results are stored in an array. Again for simplicity, it was assumed that in a multi-station configuration, all travellers using transit will depart using the central rail station and the main airport. This assumption is reasonable since the majority of American cities have hub-and-spoke type transit systems which will usually result in minimum access time if travellers departed via a city centre location. Using the freeflow speeds, distances, number of transfers required, and an assumed delay per transfer (12 minutes), the access time to the intercity terminal was calculated for each mode from each quadrant. Some example results are shown:

Transfers Required (Transit to Air)						Access Time (Transit to Air, hrs)				
1	2	2	1	0		1	1.04	0.88	0.52	0.16
2	1	1	0	1		1.2	0.84	0.68	0.32	0.52
2	1	0	1	2		1.2	0.84	0.48	0.68	0.88
2	0	1	1	2		1.2	0.64	0.84	0.84	1.04
0	2	2	2	1		0.8	1.2	1.2	1.2	1

Plate 5.7: Model Spreadsheet, Part 3

Note that transit access time along the Northeast-Southwest diagonal is particularly good, consistent with the assumption of the X-shaped trunk rail network, and an airport located at the extreme Northeast corner of the metropolitan area. Although these results are clearly for an arbitrary city, because the model captures in an impressionist fashion all the common features of a metropolitan area that are likely to drive mode choice, the results that are obtained may be generally applicable.

Assessing Transit Access Time

Using the access time matrix, it is possible to calculate the transit mode share for each cell. It is clear that the transit mode share would be different for each cell, and for each line haul mode. Since cell (1,1) lies on the subway line that serves the downtown union station, transit mode share for rail passengers from that cell would be particularly good; however, to reach the airport from the same cell, the passengers must transfer at a centre-city transfer point, resulting in lower transit mode share relative to the private auto.

The data in the analysis captures the key features of some typical transit network designs: the airport is served by one diagonal subway line, while the railroad union station (and satellite stations, if any), are served by a different diagonal subway line, and in most part of the city the population must reach both the airport and the union station by making one or more transfers, either subway-to-subway, or bus-to-subway. The features described applies in Boston, New York, Philadelphia, and Washington (although modal transfer specifics might be different, e.g. Philadelphia airport is reached only by regional rail service). These assumptions are not intended to be accurate depictions of any city in particular or of any transit system designs. Of course, with a specific city in mind, much more detailed models could be developed; however, it is hard to imagine a transit system where significantly more people (cells) would have a one-seat ride to the union station. Sensitivity analyses could show to what extent the transit mode-share changes with number of transfers. In calibrating this model, the average transit mode share across the metropolitan area for trips to the airport was assumed to be around 15% -- not unrealistic for transit-oriented cities with heavy-duty transit infrastructure and good intermodal connexions (Horowitz, 2003).

Line-haul Mode Share

Having obtained the access modal split matrix, it is then possible to calculate the line-haul mode split between the three modes under discussion. However, this has to be done for each cell, since each cell has a different access time, thus a different total trip-time, thus a different mode share. The critical concept here is that the accessibility of the line-haul terminal (be that the airport, or the rail station), affects the total intermodal journey time and thus the line-haul mode share. This reflects the argument that high-speed rail advocates have often used to promote high-speed rail -- that it brings you right to the heart of the downtown. By calculating the mode share for each cell, and integrating the demand density across the metropolitan area, it is possible to quantify the extent to which this downtown advantage plays a role in attracting traffic, and how much traffic is actually coming from the suburbs.

The most recent numbers in intercity travel surveys show a vast majority of the traffic (more than 90% in most markets) is captured by the highway mode (American Travel Survey, 1995). Thus, potentially, enhancing the access for collective carriers could apply both to the airlines and intercity rail. Indeed, airport access has already been identified as an issue constraining the growth of domestic aviation, although it is not clear that the airport access research has focused on reducing the access time to airport as an airline growth strategy (see Chapter 3), instead it has focused on moving as many people to the airport by mass-transit as possible. The argument that aviation technology is inherently constrained by the long access time, due to its massive land requirements, is not often heard, but is potentially important to the passenger rail industry.

Rail Mode shares, by cell						Air Mode shares, by cell						Auto Mode shares, by cell				
0.37	0.35	0.34	0.31	0.29		0.4	0.42	0.44	0.47	0.51		0.23	0.22	0.22	0.21	0.2
0.37	0.38	0.34	0.32	0.31		0.39	0.4	0.44	0.47	0.47		0.24	0.22	0.22	0.21	0.21
0.38	0.36	0.37	0.34	0.34		0.38	0.41	0.42	0.44	0.44		0.24	0.24	0.21	0.22	0.22
0.37	0.36	0.36	0.38	0.35		0.39	0.41	0.41	0.4	0.42		0.25	0.23	0.24	0.22	0.22
0.35	0.37	0.38	0.37	0.37		0.4	0.39	0.38	0.39	0.4		0.25	0.25	0.24	0.24	0.23

Plate 5.8: Model Spreadsheet, Part 4

While calibrating this model, it was discovered that there is an unexplained tendency for the current population to congregate towards the private auto. A likely explanation for this is that auto dominates for more than one person trips, and the majority of intercity travellers travel in parties of two or more. Although a collective transportation market share of up to 25% is possible in cities with good transit infrastructure such as New York City, the majority of intercity trips are still made by the private auto. The auto mode shares predicted by this model seem unrealistically low. There are a number of reasons for this. The model in question only considered utility of time, and did not consider two important mode-choice drivers: the actual costs of making the trip, and the perceived out-of-pocket costs of making the trip.

The distinction between actual costs and perceived costs are important: most consumers, when choosing the private auto, consider only out-of-pocket gasoline and toll costs; the ownership costs, maintenance costs, insurance costs, infrastructure costs and other costs not directly associated with the trip are attributed as 'cost of living' and thus sunk costs. When choosing a collective mode, these 'hidden' costs are charged up front by the carrier. Thus, the auto trip appear so generally attractive that in many cases consumers do not even consider other modes as a viable option. The other issue is that the consumers are making a constrained optimization decision given that the sunk costs associated with the auto (often in the form of a car loan and pre-paid insurance) is a commitment they cannot retract from. This explains the popularity of the auto in intercity markets even though rationally it seems to be a terrible choice for one individual.

5.4.3 Results and Sensitivity Analyses

The model initially predicted a 35% mode share based on the input assumptions that were made, with three stations distributed in a linear fashion within the metropolitan area. By removing two of the stations and leaving the downtown union station, the model suggested that the rail market share will fall to 33%. Significantly, with the single station, it was necessary to upgrade the average line-haul speed from 90mph to 100mph to bring the market share back to 35%, recovering the market share 'lost' when the number of stations were reduced.

The result may seem somewhat obvious, and could have been found simply by trading off average access time and line-haul time. However, many in the transportation professional community still focus on speed more than access or total travel time.

There are a number of problems with this model that preclude useful sensitivity analyses. When testing sensitivity to distance, the model gave the result expected: that better access for shorter corridors created more impact than better access for longer corridors, since the high line-haul speeds of the air service became relatively more important as the length of the corridor increased.

5.4.4 Discussion

The small changes observed in this model could easily be dismissed as noise. However, these are important noise. This model suggests that simply by opening two new access nodes, along the alignment of an existing right-of-way, could have as much impact as upgrading the entire 250-mile line to achieve an increase in average speed of 10mph. The 2% (or less) market share changes are not significant, against the backdrop of 80% or more of intercity trips completed by private auto.

The important question here is, where else in the transportation network, could you add two nodes, to achieve such dramatic impact? This is an operating environment in which hundreds of million of taxpayer dollars could be justified to re-open a commuter rail line which will carry a measly 12,000 trips per day – something like a 0.24% market share increase in daily commuting trips.

As will be examined in the next section, these extra nodes (and perhaps new urban rights-of-way) could bring important benefits for the city, if the right-of-way could be shared between urban and intercity transportation.

5.5 Analysis of Alternate Technologies

The application of Performance-Based Technology Scanning (PBTS) in this case, would pit a number of alternatives against one another. The utility resulting from shorter access time and minimization of transfers would be traded off against incremental speed improvements, amenities enhancements, and deployment of new e-commerce technologies. The methodology used is one of systems analysis: instead of focusing too closely on a specific geographic situation, with its specific nuances, the situation is simplified to give a general idea as to which areas are the most promising. Having identified the likely mix of technologies required, detailed engineering analyses for a specific corridor could then be carried out to robustly demonstrate a business case for a specific enhancement package.

Base Assumptions

The base case is a 200-mile corridor connecting major population centers, with three intermediate stops and competitive highway and air access between the metropolises. The current air service provides an in-vehicle time of 75 minutes, plus access time from out-of-town locations and extra terminal time for security clearance. The current highway access is provided by high-capacity urban expressways through the middle of all population centers en route, and the current rail access is provided by a traditional-style rail service that provides a service at an average speed of 75mph, stopping at all downtown areas. Not surprisingly, the rail carrier is finding it difficult to attract any customers. The government has mandated $2 billion to be spent on rail infrastructure in this corridor, to bring it up to standard. The question is, how should we spend the funds?

The High Speed Rail Proposal

The high speed rail advocates have tabled a proposal which would spend the $2 billion exclusively on right-of-way enhancements to bring the maximum track speed up to 150mph, a modern standard. Engineering analyses has shown that in this particular corridor, an average speed of 110mph is sustainable if all intermediate stops were removed, and $2 billion were spent on right-of-way improvements in the most cost-effective manner, including the use of tilting vehicles.

The e-train Proposal

The electronic commerce advocates and the business community have advocated a proposal which would install high-speed wireless internet service points along the right of way which would turn the train into a mobile office. Computer power-supply points would be provided in the train, and the trains would be refitted to provide space for working, including a limited number of public terminals which could be used by patrons without laptop computers. The wireless service points will also ensure high-quality, low-price cell phone calls, for those equipped with the suitable cell phone plans. Together, this package of improvements will cost $2 billion in new vehicles, telecommunications infrastructure, and other equipment. The wireless service will also create positive externalities (worth $30m annually) by providing cheaper wireless communication for the residents adjacent to the rights-of-way.

The Hotel Train Proposal

The vacation travellers have suggested that the $2 billion could go towards subsidizing luxury cruise trains between the two metropolises to bring luxury rail fares in line with current airline levels. $500 million of the funds would be spent on station improvements at either end, to institute such features as walkways to the downtown tourist centers and resorts, as well as setting up concessions in and around the stations to create a transit-center mall; another $500 million could be spent on new luxury railcars; while the $1 billion remaining would be invested, and its proceeds used as a fare-stabilization fund that will keep the fares at affordable levels.

The Downtown Proposal

The neighbourhood groups from one city has presented a plan for expending the $2 billion on a new rail alignment downtown which would move the rail alignment into a more affluent part of the city. $1 billion of the funds would be expended in creating two new Park & Ride stops and other urban neighbourhood stops that did not previously exist, while the remaining $1 billion would be expended in constructing a tunnel to by-pass a congested freight rail yard in the city, resulting in a 5-minute journey time saving. The new by-pass will also serve as a subway alignment that creates positive externalities (worth $60m annually) by providing transit in an area that did not previously have heavy-rail transit.

Benefit Analyses

Given the fixed-cost nature of this analysis, it is possible to simply consider the benefits in each of the proposals, and presumably the scheme with the most benefits is the best scheme. This does not necessarily mean that scheme should be chosen – whether the scheme is viable would depend on such matters as the opportunity cost of funds, which are outside the scope of this analysis. The application of the PBTS framework is intended to demonstrate that such completely different benefits and proposals could be evaluated with a utility analyses framework.

	Base	**HSR**	**e-train**	**Hotel**	**Downtow**
Distance	200	200	200	200	200
Speed	75	110	75	75	75
Rail Time (rt)/hrs	2.7	1.8	2.7	2.7	2.7
Access Time (at)/hrs	1	1	1	1	0.5
Disutility of rt/$-hr	20	20	15	12.5	20
Disutility of at/$-hr	30	30	30	30	30
rt saved/person	0	0.85	0	0	0
at saved/person	0	0	0	0	0.5
Annual Ridership/m	4.0	6.0	4.5	4.5	4.5
Sum(rt saved)	0	5.1	0	0	0
Sum(at saved)	0	0	0	0	2.3
Sum(utility gen.)/m	0	102	60	90	68
Externalities/m	0	0	30	0	60
Annual Benefits/m	**0**	**102**	**90**	**90**	**128**

Plate 5.9: Model 2 Spreadsheet 1

Using this very simple framework, coupled with an equally simple set of assumptions, the downtown access package was found to be the most beneficial. Interestingly, the high speed rail option was found to benefit the riders the most, while the downtown access and e-train proposals have significantly positive externalities.

Other Possible Analytic Improvements

The analysis performed here is simply an example, and contains many assumptions which may or may not be justified. It is very simplistic, but it serves to illustrate the gamut of issues that a project evaluator or public funding body must consider when contemplating investment. A private company would obviously only be interested in benefits they are able to capture; in the case of high speed rail, if rail is a price-taker in the market, perhaps none of the journey time advantages would actually be captured in the revenue stream. On the other hand, in the e-train exampline, it may be possible to capture these non-transportation benefits by offering the wireless telecommunication services to the abutters, at a fee.

To form a defensible analysis of technologies, each of the values of benefits must be substantiated either through a revenue model, or other methods of evaluating consumer surplus. As detailed in another paper, Lu & Martland (2003) investigated the cost-effectiveness of a number of high-speed rail technology options: (a) dedicated right-of-way high speed rail; (b) magnetically levitated ground transportation; (c) a number of incremental retro-fit methods, including an adaptation of maglev technologies to conventional rail lines. The study found that, on the basis of infrastructure investment cost per minute saved over typical terrain in the United States, conventional route improvements (such as minor realignments of existing rail rights of way) were about equally cost effective as the incremental maglev method, where guidance magnets are retro-fitted to existing infrastructure to enable curving at higher speeds. The benefit of incremental maglev over conventional route improvement is that it is able to achieve much more journey time reduction than small-scale realignment. Both of the approaches frequently advocated by high-speed rail lobby groups, new conventional HSR link, and new maglev link, was less than half as cost effective, even though a new maglev link could potentially achieve many more minute savings. There is thus a diminishing return effect with respect to time saved. Those who are interested are encouraged to read that paper.

Geographically-based analyses, which can be carried out by dividing metropolitan areas up into cells and calculating journey time from each cell to the nearest access point, will give us a much better idea as to how much access time can be feasibly saved. Multimodal mode-spilt analysis will tell us how much the access is likely to affect existing highway and airline patrons. The positive externalities could be subject to a much more rigourous analyses than is presented here. The utility of time and of time-saved could be disaggregated into different market segments and other such detail. For instance, in Slide 5.4, we demonstrated that an average time saving of 20 minutes per passenger trip was possible if London rail terminals were directly connected with each other.

Computer simulations could be used to calculate the loci of influence of introducing a new rail terminal, as detailed in Slide 12 in Appendix B. The mode-shares were predicted using a simple nested logit model, involving two choice phases: the mode-choice for the access mode, that subsequently affects the mode-choice of the line-haul mode by altering the total trip time that includes the access time. By virtue of the short total trip time, airlines can dominate an entire metropolitan area from one single airport, unless rail access in a locality is extremely good. This suggests that introducing extra stations in the metropolitan area will have a very positive benefit for the rail carrier.

The Track Capacity Studies, shown in Slides 9 and 25 in Appendix B, used an operations-planning simulation model implemented in Excel to calculate the likely positive externalities by introducing an intercity rail tunnel through the downtown and some city neighbourhoods. The operations planning simulation showed that track sharing is indeed feasible given operating discipline, while the ridership model showed that, especially on low-volume intercity corridors, the most significant benefit of an intercity tunnel is through carrying the city population while the tunnel isn't being used by an intercity train. The high-speed rail option in the present analyses assumed fairly high ridership volumes of 4 million annually; if that number were lower, the external benefit of the downtown upgrade begins to look much more significant against the rider benefits of having a faster train.

5.6 Conclusions

In this chapter, we have demonstrated an important methodology with which high-speed rail schemes should be evaluated. Firstly, the issue of access to high speed rail services has to be taken seriously, if high-speed rail advocates wishes to be considered as a viable voice in promoting mass intercity transportation. Secondly, to demonstrate that a high speed is really necessary, the speed enhancement must be explicitly traded off against other possible enhancements in a utility framework.

References

Martland, Carl D., Lu, Alex et al. Performance Based Technology Scanning Applied to Downtown Access, *Performance Based Technology Scanning Applied to Passenger Rail*, MIT/UIC Working Paper WP-2002-02, Massachusetts Institute of Technology, Cambridge, Mass. (2002)

Lu, Lexcie., Martland, Carl D. et al. Performance Based Technology Scanning for Intercity Passenger Rail Systems: The Incremental Maglev and Railroad Maglevication as an Option for Ultra High Speed Rail. *Proceedings of 2003 IGERT Advanced Research Conference*, University of California, Davis, Calif. (2003)

Lu, Lexcie. From the Zephyrs and the Limiteds to the 21st Century MetroFlyer. Presentation to the Intercity Rail Financing Subcommittee, Transportation Research Board.. MIT/UIC Web Document No. xxxxx, Cambridge, Mass. (2002).

Chapter 6

High Speed Rail Route Planning for Overnight Services

This chapter reviews a Swedish study on overnight train service (Troche, 1999), and our own study of overnight train service, carried out independently in 2002. Our study is roughly based on the Performance-Based Technology Scanning framework, while the Swedish study came from a carrier operations planning perspective. The central thesis is that, given the general principles of: (1) shared traffic rights-of-way, (2) market-oriented service planning, (3) no 'subsidized' competition, in certain markets overnight rail service can actually be a lucrative proposition. Corridor proponents often incorrectly see overnight service as 'the enemy'; whereas the two services actually complement each other, and both have their own strengths and problems.

The main problem with corridor service is the service frequency that is needed to compete with other transportation options such as the private auto, highway, air, or even buses, which carry much smaller number of passengers per vehicle-formation. In very high density corridors, high-speed, high-frequency, short-distance service has carved out a niche; however, to support that type of service, great demand density is required. In longer and lighter density corridors, rail service should exploit a different niche -- a service that operates a few times daily overnight, offering a comfortable journey and the opportunity to wake up at the destination with a full business day ahead. While overnight services have higher vehicle and infrastructure costs, they requires much lower frequency and are likely to carry far higher load factors due to temporal-consolidation. Operationally speaking, it is somewhat more cumbersome and must be managed in a different way – however, efficiently run, the economics should work in at least a number of corridors.

6.1 Review and Discussion of the Swedish Findings

Gerhard Troche has done the intercity rail community a service by issuing his working paper on overnight services. He addresses issues that are poorly understood by those who have little experience with night-train traffic -- which is a strange combination that falls somewhere between the expertise of the freight transportation, regional rail, and hotel management communities. The overnight service is like a freight train because it is a 24-hour operation and typically uses loose cars; it is like a regional rail because it transports people and is important for it to arrive on time; it is like a hotel because amenities and service are a very important part of the package. Thus, the issues are: (1) railcar interior design; (2) trunk service design; (3) connection with day trains. These may sound like familiar issues, but they require a different kind of thinking than those in the parts of the transportation industry who are seeking to minimize journey times are used to. Unfortunately, with the demise of overnight ferry services in Europe, and the demise of the streamliners and much of the rail cruise industry in North America, much of this expertise has been lost.

Plate 6.1: Overnight Services do not have to rely on fully depreciated assets

Economic Issues

Troche paints a very negative but realistic picture of night-train economics. Intercity buses, airlines, and the private auto have all become much more formidable competitor since even the 1980s. An issue unique to Europe is the added complication of the need for cross-border cooperation. Due to the generally increasing speeds of rail services, overnight trains are travelling over increasingly long distances and thus have a higher probability of crossing national boundaries. Troche asserts that even though the day-train market had been growing and night-train market shrinking, this is because there had been major investment in day-trains but no comprehensive effort to coordinate night-train traffic. Despite the increasing prominence of the day train, the night-train is not 'just a niche market' -- it is an important area of the passenger rail core business, especially if the full network benefits of multi-national high-speed networks are to be realized.

Troche asserts the night train suffers from high costs and poor utilization. A simple analysis of available-seat/bed-miles (ASM) per vehicle shows the regional day train to be four times as productive as the night train, and the high-speed train to be six times as productive. In addition, there are additional crew costs associated with the night train. These are accurate characterizations, except that in regions of low day-train demand, analysis of revenue-passenger-miles (RPM) will tell a different story. Using Boeing's Decision Window Model, the following temporal-demand prediction for an eight-hour trip is obtained:

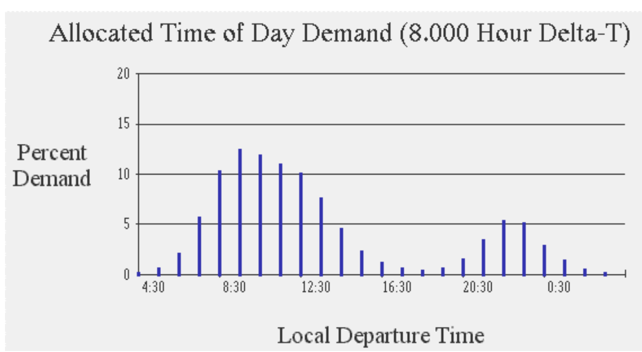


Plate 6.2 Allocated Time of Day Demand (8.000 Hour Delta-T)

This demand curve suggests that for an eight-hour trip, the majority of the people would prefer to depart some time between 7am and 1pm, arriving between 3pm and 9pm, although there is a sizable minority who would choose to catch the 'red-eye' and arrive the next morning. A few caveats are important here: (1) this demand curve assumes airline passengers, who had no other ways of getting there; (2) this demand curve assumes the effect of journey time is unimportant: the time-of-day demand being driven entirely from the path with the shortest journey time, thus demand allocation sensitive to both journey time and departure time is not possible to capture with this model.

Using Boeing's model, six paths were created: one was the overnight train, leaving at 10pm and arriving at 6am with AEM-type equipment. The remainder are bihourly express trains (that also take eight hours), using Acela-type equipment. This timetable is typical of Northeast Corridor services between Boston, Massachusetts and Washington, D.C. The model shows that that the night-train captures 16.4% of the traffic with 16.7% of the resources (one out of six trainsets). Interestingly, the *Morning Congressional* leaving at 6am to arrive at 2pm actually only captures 8.7% of the traffic!

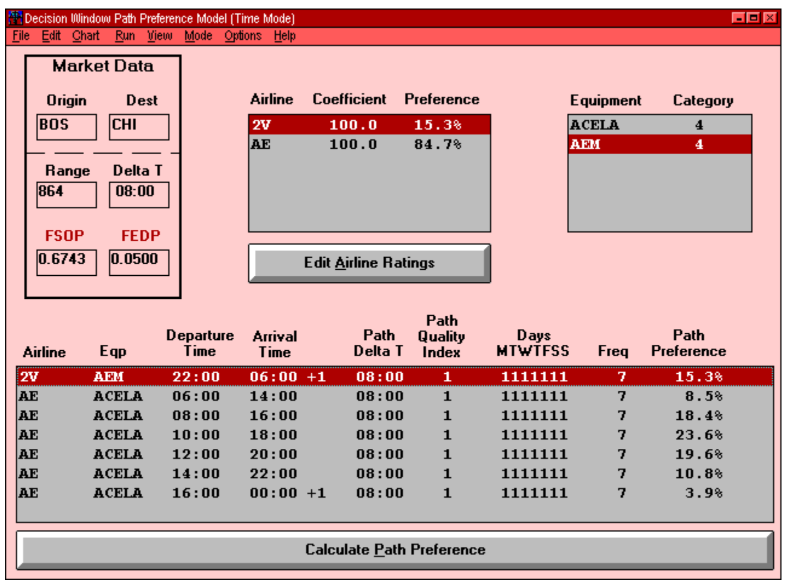


Plate 6.3 Decision Window Path Preference Model (Time Model)

This is obviously an overly simplistic assessment of the situation. However, it is clear from this example that corridor daytime trains can be resource heavy because of the requirement for frequent service. If service went from bihourly to hourly, the night-train turns out to be the most effective at capturing passengers (14.2%), using 10% of the resources (one trainset out of ten) -- more than the maximum of 13.4% captured by the 9am departure and 13.2% captured by the 10am departure.

The other issue not considered here is the effect of airline competition. As previously mentioned, the demand curve is for all passengers. In corridors where airlines are operating, the temporal-demands are actually split by the carriers based on both journey time and time of arrival and departure. The demand curve are likely to be driven at least partially by the equivalent 2-hour journey available via air. The shape of the curve is likely to depend on complex factors such as purpose of the trip, and itinerary planning. Day-trippers who can do so are likely to fly both ways; day-trippers or vacation-makers preferring no loss of sleep are likely to travel one-way by overnight train and fly the other way. The temporal-demand is at least partially driven by the choice of mode. Given an institutional framework that allows free-choice between overnight train and flights, with no financial penalties and any combination of air and overnight train, many more people may choose to take the one-way overnight train.

The problem for overnight trains with the current set up is that travellers often are constrained to choosing either a round-trip airfare or a round-trip train fare. The key innovation here is for the computer reservation systems (CRS) algorithms to smarten-up when planning itineraries: when pricing a round-trip between night-train served cities, rail-air combination ought to show up alongside day-trip flight combinations. Willingness-to-pay pricing could then be applied at this point, extracting the same consumer surplus, whatever combination of modes the passenger happen to choose.

None of this work is intended to discredit the current Amtrak timetable. The reason for the 6am Acela Express departure from Boston is related to day-tripping New York passengers, not those wishing to arrive in Washington between 1pm and 2pm. In any case, the Boston to New York passengers give rise to a much higher yield than the longer-distance passengers. The additional complication with high-speed trains is that, unlike airlines, trains stop en-route thus a train-path is not necessarily the same as a passenger-path.

The day-train appears much more lucrative and productive as the schedule is designed for the day train vehicle to criss-cross between regions of very high demand density, compared to the market that night-trains operate in. If the day-train vehicles were used to offer locals along the route of the night-train, it is likely that the night-train is likely to achieve much higher productivity and load factor. The core issue here is that the 600-mile market has always been a much less lucrative market for rail -- and the key is to time night-trains such that arrivals can also serve peak daytime demand as the trains travel into town early in the morning. The creative tweaking approach is needed in such markets; abandoning such a market may be a management response to focusing on the most leveraged market segment.

There is a research need for an integrated intercity demand model which is sensitive to mode-preferences as well as schedule-preferences. It is conceivable that with shared right-of-way with other corridor trains and freight trains, the overall system costs could be reduced by carrying some of the current airline passengers who are flying in the morning peak in overnight trains instead, at the same time improving the passenger utility by allowing passengers to avoid waking up early for a flight. Part of the problem is that airlines and railroads have never cooperated to this extent, and airlines often don't understand that the highest fixed costs are incurred during the peak hour. The integrated system may well become impossible to manage due to its complex nature involving many modes, but the idea represent an as-yet untrialled transportation system configuration.

Further Investigation of Time-of-Day Demand Curves

Using Boeing's Decision Window Model, an attempt was made to assess the amount of traffic that could be captured with an overnight train that is equivalent to a two-hour flight either in the morning or in the late evening. Using airline demand data for two-hour air journeys, it was found that 79% of demand occurs between arrival times of 11am and 10pm, a time-period when overnight trains are not competitive. The demand curve showed three peaks: a morning peak, presumably from business travellers travelling to a mid-morning meeting; a midday peak, presumably predominantly from vacation travellers, and an evening peak, presumably from business travellers heading homewards after work or meeting.

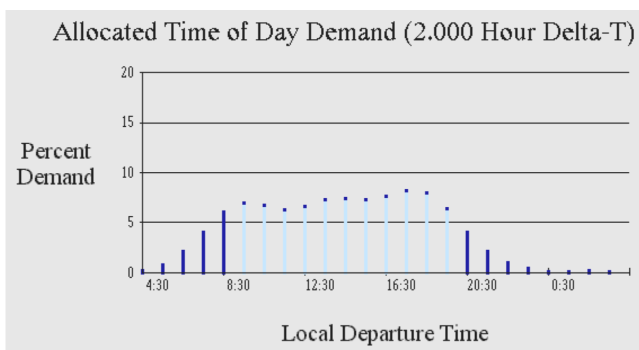


Plate 6.4 Allocated Time of Day Demand (2.000 Hour Delta-T)

It is not clear to what extent this demand curve, which is based entirely on airline passenger surveys, would be affected by the availability of overnight services. Using a simple model developed in Excel, a hybrid demand curve was created from the temporal-demand prediction for an eight-hour trip, and that from a two-hour trip, based on considerations of overnight services competitiveness for demands that occur within a specific departure time window. Based on this line of reasoning, the overnight services competitiveness depends on one key variable: to what extent overnight services are able to capture the airline demands that occur during the periods when airline services and overnight services are reasonably substitutable? Consequently, this key variable also affects to what extent the demand curve would be affected by the overnight services: if overnight services can capture most of the traffic, then the demand curve would have a much more substantial late-evening peak centered around 10pm; if overnight services captures very little traffic, the demand curve would be very similar to the one shown above for a two-hour trip. Hereafter, this variable would be referred to as the overnight services base potential (ONSBP). Using ONSBP of 50%, the following time-of-day demand curve was obtained:

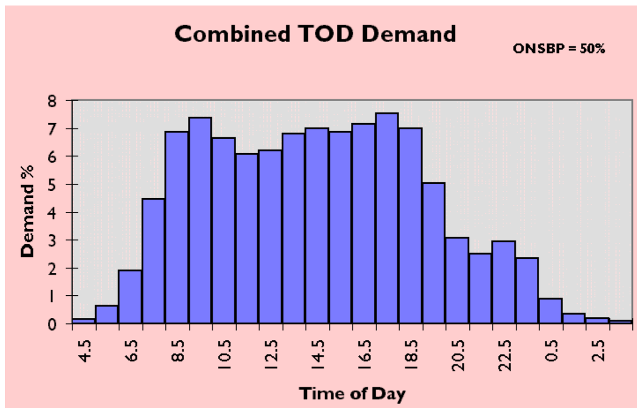


Plate 6.5 Combined TOD Demand, OSNBP = 50%

From this graph, it is fairly clear that overnight services would not have a major impact on the time-of-day demand curve -- since ONSBP of 50% is perhaps unrealistically high, at least for a market that isn't particularly mature. Based on the British Rail experience, where in corridors such as Glasgow-London, overnight services captured something like a 6% market share of daytime trains (where daytime trains achieved a roughly 50% market share of total traffic), ONSBP really should be something more like 3%! However, these numbers can be misleading, since overnight services in the Glasgow-London corridor is not well-developed and certainly British Rail's continual active effort to kill it between 1980 and 1994 has contributed to the low market share. As Troche notes in his report, overnight services have never particularly suffered from the lack of demand -- it's the high operating costs that is the problem! As previously discussed, the apparently high operating costs may have resulted from corridors that are not long enough to justify overnight services.

Thus, in a corridor where overnight services can potentially be justified (a market like New York to Chicago), we can perhaps expect overnight services to capture about 20% of market share available during the target periods. Sensitivity analyses gave the following results:

Sensitivity Analyses, ONSBP v.s. Rail Market Share						
Lexcie Lu, MIT CTS, 19 April, 2003						
Daily Passengers in Corridor			4,000			
Train Capacity (8 cars)			200			
ONSBP%	0%	3%	6%	10%	20%	50%
Rail MS	0.00%	0.71%	1.44%	2.4%	5.0%	13.8%
Pax/day	0	28.4	57.6	97.6	201.6	550.4
Trainloads	0	0.14	0.29	0.49	1.0	2.8

Plate 6.6 Sensitivity Analyses, OSNBP v.s. Rail Market Share

From this study, it is evident that the breakeven point for a large market requires an ONSBP of about 20% -- not unachievable, but not easy. For instance, in Italy, where the overnight services network is very developed, 43% of all travel takes place during the night (Troche, 1999). The initial study (Lu & Martland, 2001) overestimates the potential demand in overnight capacity, as it did not consider the effect of portions of trips that occur during the time window when overnight services are competitive. The current study suggests that daily passengers volumes of at least 4,000 and an OSNBP of 20% is required for operation.

However, this is not to suggest that overnight services are not an important product. With ONSBP of 20%, the largest markets could support at least one train daily; smaller markets could be consolidated, since overnight train is able to "pick up" in many cities before embarking on the long overnight journey, a point discussed in more detail in the Swedish report.

Given the above quantitative analysis, it appears that the many advantages of overnight train are overshadowed by the fact that air travel is a lot faster, and could take place during the day without too much intrusion into the traveller's schedule in the 800~1,200 mile market. Where overnight services are available, and effectively marketed in conjunction with air service, overnight services may offer an attractive alternative to travellers wishing to make their journey time "disappear" while they sleep. How the market will respond to this kind of trip-based multi-modal marketing is anybody's guess, as airlines around the world have never marketed overnight rail services as an alternative to early morning or late evening flights in an integrated sort of way.

6.2 Review and Discussion of the Initial Study

The initial study (Lu & Martland, 2001) was primarily based on the author's experience while working with a train operating company responsible for operating the Anglo-Scottish Sleeper service in Britain.

Previously under British Rail, Britain boasted one of the most extensive overnight train services anywhere in the world, with services resembling a totally-connected network between most important origins and destinations. During the sectorization in the late 1980s, overnight services were singled out as a business sector that generated heavy losses. As train speeds increase on day time trains, journey time between most major population centers broke the four-hour psychological barrier, resulting in heavy cut-back of overnight services. As of 2000, only Anglo-Scottish overnight services remained, and overnight service to the north of Scotland was continually under threat.

The initial study examined ways in which the overnight service concept could be made profitable in North America, where urban centers are much further apart compared to those in Britain, and high-speed daytime train infrastructure much more lacking. As part of the study, many of the issues surrounding the overnight rail service design and infrastructure cost sharing with proposed high-speed daytime corridors were explored. In this section, the results of the initial study are compared and contrasted with those from the earlier Swedish study, which is predominantly aimed towards connection from Scandinavia to continental Europe. Potentially, the same findings could apply to connections from Britain to continental Europe, as the Scandinavian peninsula bores much resemblance to the British Isles in terms of their relative location from the major European population centers.

In the initial study, the Capitol Flyer scheme is suggested as a strategic solution to North America's high-speed rail problems. Local high-speed corridors could be planned in such a way as to accommodate long-distance overnight services, which provide an alternative to airlines both as redundancy and to provide travel options. In developing the initial study, focus of attention is devoted to capital costs as rebuilding the passenger rail infrastructure at vast capital expense would be necessary before either high-speed corridor or overnight rail would become competitive.

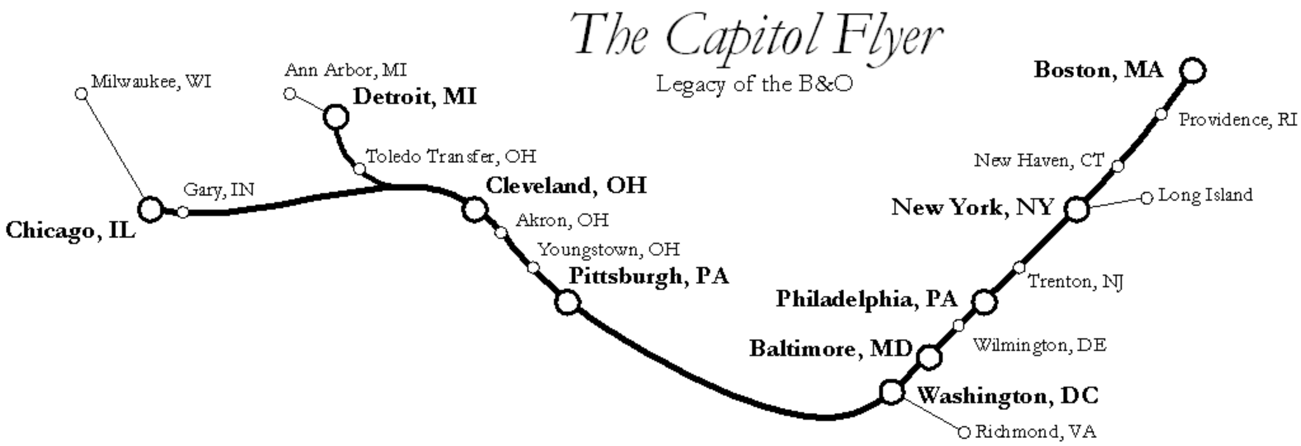


Plate 6.7 The Capitol Flyer – Legacy of the B&O
(and successful overnight high speed rail proposal – by joining existing corridors)

<u>Market Segmentation</u>

The Swedish makes a very important observation: that night-trains must be rolling hotels and youth hostels at the same time (Troche, 1999). In the airline industry, market-segmentation has been achieved through restrictions on ticket flexibility -- with the results that airlines were able to build a network that served both business and leisure passengers. The airline model works, as long as there are no low-price entrants. The low-price entrant builds a smaller network serving only the most profitable flows, resulting in the loss of cross-subsidy between routes and service levels. In the overnight rail market, market segmentation through ticket restrictions can be difficult to achieve since the traffic volumes are low, and the advantage of the overnight service (i.e. avoidance of an early morning or late evening flight) is insufficient to extract high willingness-to-pay from high-margin market segments.

The strategy suggested earlier of introducing multimodal, trip-based pricing may solve some of these problems. With control over both the rail and airline links, the carrier is able to price in a way which extracts most of the consumer surplus; in essence, the overnight rail will no longer compete directly with the early morning flight, but instead the price is simply based on competitiveness of other airlines, the overnight bus, and the private auto. Given the consumer's typical utility function, it is likely that overnight rail would be used to an extent much greater than the status quo if the competing airline trips were priced differently.

As it stands, it is not clear that either airline, rail, or highway costs reflect a fully-allocated costs of providing transportation; both air and highway modes have Federal capital subsidy, and increasingly rail carriers are accepting infrastructure subsidies from the Federal government. Permitting joint-pricing (or collusion) between air and rail carriers may actually induce more rational choices on the part of the consumer as to how they plan their trip. As demonstrated in the last section and concluded in the Swedish study, the lack of demand is not much of a problem in the overnight market; while the overnight rail costs are more than an early morning flight for the same travel distance, it also generates much more utility. If distorted market conditions in the short-run did not exist in the airline industry (i.e. infrastructure subsidies, sunk cost in expensive fleet), and costs were fully allocated, overnight rail could fit better with many consumers' travel schedules and plans.

In short, subsidy of air and highway travel has led to over-consumption of these modes that cannot be rectified in the short term due to the problems of long-lasting assets. As a result, from a strictly utility point of view, consumers are consuming less overnight rail travel (and more airline travel) than is optimal. Thus, market segmentation, airline-style, is not possible, as overnight rail has no pricing power.

Service Planning Issues

The Swedish study distinguished between three types of night-train markets: (1) evening departure, morning arrival; (2) departure after work, morning arrival; (3) departure after work, late morning arrival. Each type of night train service also served markets of a particular distance (although the distance should be regarded as an upper bound, as it is always possible to operate the trains at a lower speed to enable a shorter distance to be covered by an overnight service. In terms of service planning, Troche observed that night trains are most competitive where a population center remains relatively isolated from a number of others (e.g., there is at least 625 miles between Stockholm and most large European cities). In those circumstances, the night-train could be an attractive alternative to airline service where direct service to most cities could only be sustained a few times daily.

Applying the concept to North America, it is conceivable that with trains operating at 90 mph, Washington is a city where this idea could perhaps be successfully applied. At 90mph, a number of important population centers are reachable overnight: Boston, Pittsburgh, Cleveland, Buffalo, Cincinnati, Atlanta, and the Carolina coast. A similar concept could perhaps be applied to Chicago. However, part of the problem is that with smaller number of population centers and a more dispersed pattern of economic activities, flights with regional jets or simply the private auto may remain a cheaper way to achieve roughly the same objectives. Because of the high capacity of an overnight train consist, it is likely to replace some four to eight daily flights, while many cities can only justify one to two flights daily to each destination.

In terms of operating principles, Troche discussed the concept of the CityNightLine philosophy. Night train would make a series of stops between 8pm and midnight, and travel non-stop between midnight and about 6am, before dropping passengers off until perhaps as late as 10am. The concept has been in practice on the Deutsche Bundersbahn network for a number of years. The concept was independently explored in the initial study (Martland & Lu, 2002), using Scotrail's Sleeper services and a hypothetical American service as examples.

Troche explored in detail the concept of designing routes with many detours, making the important observation that while journey time is an important drive in day traffic, in night traffic the arrival and departure times are more critical. As a result, day trains often have a trunk section with many connecting branches, while overnight trains can take circuitous routes and 'drop-off' long-distance passengers in a polycentric region like a local.

In the initial study, this concept was applied to North America geography, while attempting to minimize costs of right-of-way since high-quality, high-speed passenger railroad network is not readily available in North America. The initial study introduced a novel way of serving long corridors such as the Northeast Corridor with overnight services. Instead of having a single train which must pick-up from a range of cities along a corridor between 8pm and midnight, two trains could be used to cover a longer corridor. Thus the first section of the train would leave from the northern end of the corridor, while the second section would leave from the middle at the same time, thus shortening the pick-up window. Cars could then be exchanged as the southern portion stops and wait for the northern portion to arrive at a predetermined location. In the morning, the two trains could head off to different destinations (or different sections of the same long corridor) to perform drop-offs. (See discussion on Cumberland Sleeping Sidings, Lu & Martland, 2002). This method of operations can be somewhat similar to today's package-express freight operations. The crucial feature of this operating plan is that the customers are not disturbed by the need to change trains in the middle of the night.

Railcar Interior Design

This topic is discussed extensive in the Swedish report. Troche has collected a comprehensive set documentation of existing state-of-practice and new ideas in terms of how to lay out an overnight railcar. His recommendations centred on a design that would combine both daytime and overnight traffic in the same vehicle, so that effective use could be made of the vehicle during the day.

Given the comprehensive coverage in the Swedish report, this topic will not be covered in this thesis. It is conceivable that more effective use of space could be made than the current Viewliner and Superliner designs, but given the paucity of intercity passenger services currently in North America, it is unlikely that railcar interior design would be a driving factor in making North American overnight rail services more viable.

<u>Other Issues in Overnight Services</u>

Attention was drawn to the fact that market expectations of transportation services in Europe may be markedly different from that in North America. In particular, some issues that have been raised with respect to adapting a proposed overnight service to the North American environment:

- *Americans like their private space more than Europeans*: One issue that has been raised in the Swedish report is the poor utilization of space in an overnight services car. If the seating density could be increased, the cost of overnight services would fall dramatically. The problem here is that to increase the seating density will necessarily require some sharing of space by fellow passengers, who may be total strangers. In Russia, overnight cars exist in which a single large compartment is shared by six travellers, none of who may have met previously. As an European, I do not find this strange, although apparently in the American culture such proposals are considered unsafe and unacceptable to most people. This may also explain in part the American's obsession with the private car, which offers a private space during travel not offered by mass transit. Certainly, security and privacy concerns are cited by non-transit users as reasons for not using transit.

- *Americans like larger space than Europeans*: Staying in a hotel may be preferable to staying on an overnight train simply because the hotel offers more space. In Europe, due to a number of practices that have been in place for a number of years, including traditionally higher costs of transportation and energy and more conservative zoning practices, the population are much more used to living within smaller spaces, which resembles a train more than it resembles a hotel room.

- *Americans wake up earlier*: In Europe, the business day starts at around 9am, although the practice varies from region to region and some regions of Spain has business schedules that calls for a midday break. In North America, because of the traditionally longer commute and earlier start of business day in some regions (8.30am in New England, 8am in Chicago), the attraction of the overnight train may be reduced because the barrier to waking up early for an insanely early flight is reduced. Although on a macro level, this should not make any difference, the author's experience suggests that business practices in Europe tended to rely much more on arriving at around 10am if intercity travel is involved, whereas in North America, because of the larger geographic area, people are much more used to scheduling later meetings. Thus, comparatively, the attraction of overnight services are eroded.

6.3 Conclusions

In this chapter, a novel way to ascertain demand in the overnight rail market is demonstrated, with inconclusive results. Although it was not possible to show that a certain level of demand for overnight services exist, it was shown that the level of demand is not zero and that a significant consumers ought to prefer overnight services if cost is not an issue and that the service was marketed jointly with air services.

The overwhelming conclusion of this investigation that the problem facing overnight services is the current institutional structure which puts air carriers at odds with rail carriers, and commuter, corridor, and freight rail carriers at odds with long-distance rail carriers. Because commuter services affect a large number of people, while air services have up to recently been considered for-profit propositions, there has not been any real effort to integrate air and overnight rail services, while efforts to improve it has been sparse compared to the amount of attention and public funds devoted to commuter rail and other travel options, such as highways.

An important point to note is that expertise in overnight services management still exists at least in one segment of the professional community. Regions (such as Sweden and Scotland) that have traditionally been isolated from major population centres have developed considerable expertise in operating overnight services and making them work. While overnight service will probably never become the most leveraged portion of rail operations, if implemented wisely and coupled with existing high-speed corridor infrastructure, it could be a much more formidable competitor than previous experience and studies have demonstrated.

Overnight rail is not an inherent loser. Overnight services possesses inherent advantages that cannot be matched by either daytime rail services, the private auto, or the airlines. It is up to managers of overnight services to exploit these advantages. Management of overnight passenger services is a separate discipline, distinct from management of daytime corridor rail services, freight services, or hotel business. Recognizing this unique position of overnight services management, railroads that have them should be establishing overnight services as a separate business sector and instituting a separate management team for such services that have special requirements and represent an unique facet of passenger rail operations.

Consumers may be willing to pay the increased costs associated with overnight services if true marginal pricing is instituted for all modes (or indeed, fully-allocated costing is instituted for all modes). There is currently no data to support or refute the claim of overnight services' poor economic performance, as the current poor economic performance of overnight service could be explained by other factors such as less-than full cost allocation on competing modes, or simply a lack of strategic management on the parts of some overnight services operators.

References

Sussman, Joseph M. and Martland Carl D. et al. Performance Based Technology Scanning Applied to Overnight Services, *Performance Based Technology Scanning Applied to Passenger Rail*, MIT/UIC Working Paper WP-2002-02, Massachusetts Institute of Technology, Cambridge, Mass. (2002)

Troche, Gerhard. Efficient Night-Train Traffic -- Problems and Prospects. Working Paper KTH/IP/AR-99/77-SE, Department of Infrastructure and Planning, Royal Institute of Technology, Sweden (1999).

[This page is intentionally left blank]

Chapter 7

Summary and Conclusions

Metropolitan access in today's large and sprawling metropolitan areas plays a vital role in rail transportation – both in terms of its competitiveness against other modes, and in terms of cost-effectiveness of investment alternatives in delivering customer satisfaction. This is generally true in all the market segments analyzed: commuter rail, regional rail, intercity rail, and overnight services.

The objective in intercity transportation is to deliver customers from their actual origin to their ultimate destination – and not from the nearest intercity transportation facility (such as an airport or a downtown union station) to the facility nearest to where they are going. Although a proportion of intercity trips are likely to originate from the downtown core, many more intercity trips today are likely to be originating from the suburbs or city neighborhoods and destined for the suburbs or city neighborhoods in a different metropolitan area. The designer of the intercity transportation system must consider both types of demands when designing a system.

This result has important implications in development of next generation high-speed rail technologies. The need to make multiple stops en-route mean that the ability of rail vehicles to accelerate from a standing stop would be an important technical attribute – an equally important one to the vehicle's ability to sustain long periods of high speed running. Many high-speed rail proposals (whether technological or project-oriented) have focused on constructing a beeline from one city center to another city center to provide the fastest point-to-point journey time. In fact, the vehicle's ability to negotiate 'classic' curvaceous infrastructure is likely to be as important if not more so than the need for sustained high speeds on unconstrained infrastructure. The reality of today's urban planning is that unconstrained infrastructure is almost impossible to retrofit, and even if it were possible, it would not serve the largest demand generators – ones that are dispersed throughout the gentrified inner-city neighborhoods. Such neighborhoods are reached through classic infrastructure that led to their development in the first place, and do not usually lie along a perfect straight line from the city center.

Just to reiterate this important concept, missed by many high-speed rail enthusiasts: the integral of demands throughout the city neighborhoods far exceeds the demand from the downtown core, even if the downtown core remains the busiest node on the whole line. High-speed approaches that calls for a hub-and-spoke network based on a downtown hub through interconnection with the municipal transit system cannot be successful if the access time to the downtown hub is so abysmal as to rule out intercity rail as a viable option against driving directly to the destination.

The types of technology that are likely to be successful in future intercity rail markets are backwards compatible. Backwards compatibility allows the new technology to be deployed incrementally, avoiding high up-front capital costs and preserving network effects. It also allows sharing of high infrastructure costs with other modes such as commuter rail and urban transit. Retrofits of existing 'classic' corridors are likely to generate less objections and disruption to existing urban fabric. Technologies that build upon current steel-wheels-on-steel-rails guidance while enhancing its performance (particularly its performance on constrained 'classic' infrastructure) – such as flange lubrication to reduce L/V ratios, and magnetic guidance to allow higher superelevation (cant deficiencies) through twisting curves, will be far more important, cost-effective, and leveraged than proposals to construct brand-new rights of way. Magnetic levitation technology could be used in an incremental fashion by providing guidance alongside, and not instead of, steel-wheels-and-steel-rails. These hybrid concepts leverage the inherent value in the existing infrastructure, and is much more important than 'shiny-go-faster' approaches that focusing on relaxing the constraints through sledgehammer-like, environmentally insensitive engineering.

These findings have important implications for transit properties around the country. Transit property should realize that sharing infrastructure costs with intercity carriers could substantially reduce the cost of providing transit. Instead of seeing the intercity operator as the 'competition' and requiring them to interchange with the local carrier only at designated points, the transit property ought to see the development of a new intercity corridor as an opportunity to provide transit service to another section of the city. Intercity and transit providers should work together to allow new intercity infrastructure to create new transit corridors. New corridors should not be constructed alongside existing corridors, but instead should be constructed such that easy interchange is accomplished with radial lines running in all directions. Not only will that design provide maximum connectivity, it will also enable the intercity carrier to tap into important business centers all over the metropolitan area and not just the downtown. The new corridor could also serve as a crosstown transit corridor.

Not one solution would work for every city. Depending on the location and condition of the existing transit and intercity infrastructure, and the economic geography of the metropolitan area, different layouts of metropolitan access infrastructure (or if you prefer, terminal district infrastructure) would be necessary. Above all, transit properties and intercity carriers need to demonstrate a willingness to work with one another in generating a good design from the customer's perspective – regardless of who will carry the passenger. There should never be a single 'interchange point' where one carrier's responsibility ends and another takes over. The issues such as: who will pay for the infrastructure, who will manage its construction, and who will own the tracks, are independent from how one might design a good transportation network. Once the network has been designed to give the best possible customer utility for the largest number of origin-destination pairs, those institutional issues would likely be resolved more easily. Links that carry predominantly intercity passengers would be paid for and operated by the intercity carrier; links that are common facilities would come under joint control, and perhaps even shared ownership – just as the union station brought together bitter rivals in 19th century railroading, the union terminal district network should bring the transit property, the intercity carriers, and perhaps even the freight carriers together to design a layout that will work for everyone.

7.1 Analysis for Specific Market Segments

7.1.1 Commuter Rail

Commuter rail represents a market segment in which customers make trips daily – and decision about commuter rail ridership generally represent a long-term commitment. Demand pattern is generally concentrated in journey-to-work trips, comprising in the morning rush of pick-up in low-density suburbs and drop-off in a high-density downtown business district. However, in today's dispersed cities, the historical downtown is no longer the only location of concentrated business activity: firstly, downtown business districts have expanded into adjacent neighborhoods and can sometimes can be as large as three miles long; secondly, some inner city neighborhoods may have gentrified and represent significant business activity.

One design of commuter rail system would call for an easy transfer from a central dispersal point to the municipal subway, sometimes necessitating 'backtracking' to get to the traveler's eventual destination. Another approach allows all incoming commuter trains from any direction to call at multiple number of 'union terminals', enabling walk access to most centers of business activity, at the expense of perhaps not following the most direct route from the suburbs to the downtown. When transfer time, terminal time and buffer time are fully considered, most travelers would prefer the latter approach. In addition, by providing a one-seat ride, the value-of-time on-board the commuter train is increased – which can translate either into higher willingness to pay for the service, or higher social benefits.

There is much empricial results supporting this analysis in current heavy-rail transit experience: more recent systems such as Bay Area Rapid Transit (BART), the Massachusetts Bay Transportation Authority (MBTA) Red Line, Metropolitan Atlanta Rapid Transit Authority (MARTA) that have a commuter-rail like character have matched their station spacing to demand density, and designed their system such that many parts of the downtown core could be reached directly. These systems would probably not have been such a runaway success if a transfer to a bus or a streetcar had been required to reach the outer fringes of the downtown after an incoming ride from the suburbs.

Plate 7.1: Boston's Commuter Rail serves both the commuter rail and regional rail functions, but the two downtown terminals make cross-metro area journeys difficult.

Access time at the downtown end of the trip is particularly critical. Suburban commuting decisions are usually driven by long-term choice of residential location and not employment. Frequently, one household may have more than one commuter rail rider who is able to share an auto at the suburban end of the trip but not the downtown end of the trip. Requiring suburban commuters to make a second transfer onto the subway before walking to their final destination could make other arrangements such as carpool drop-off (either at place of work or at a Kiss-n'-Ride transit terminal) much more attractive, and would diminish both the social benefit and the political support for commuter rail. Commuter rail ought to be seen as a limited-stop subway for auto-welding suburbanites who wish to reach a variety of locations in the city – just as an urbanite could travel from an urban neighbourhood to a variety of destinations served by a trunk line subway. The difference is that a suburbanite would drive to their transit access location while an urbanite would walk or take a bus. The 'commuter rail terminates at union station and serves the center of downtown travelers' paradigm is no longer tenable in all but the most compact and dense city centers today.

7.1.2 Regional Rail

The definition of regional rail is not always clear. For the purposes of this discussion, regional rail could be one of two types of services: (1) longer distance exurban commuting services which are unsuitable for daily use, but generally allows day-trips by business travelers – examples would include the Keystone and Empire corridors in the Northeast; (2) services that transcend the metropolitan area from one suburb to another, achieved by either a connection or a run-through train in the downtown – for this type of journey, the commuter rail schedule is often so constraining that daily commute is not a possibility except for the most determined employee. This definition allows regional rail to describe a different market from commuter rail, even if travelers from both markets may travel on the same physical trains. The needs of the two types of customers are very different, and serve to illustrate the importance of considering all customers in systems design.

Because of the lower demand density at both origin and destination for the crosstown regional rail rider, it is a practical impossibility to provide convenient access through the rail mode. However, it is likely that the leisure travelers in this market will be able to arrange a pick-up, as the motivation for travel is often to visit a friend in a different suburb. The need for regional rail customers is the ability to transfer easily once they reach the downtown.

Although the 'union station' design could potentially accomplish an easy transfer, several practical problems generally prevents that arrangement from being satisfactory. Firstly, the existing infrastructure in large cities is usually such that there are more than one union terminals, constructed in the late 19th century by different railroads. For crosstown travel, frequently a short transfer in a taxicab or the urban transit system is required. This two-transfer solution is a poor alternative compared to driving, and is unlikely to create the type of competitive atmosphere where regional rail is considered an alternative other than for the autoless suburban poor. Secondly, there are usually good operational reasons for central stations to be decentralized in a large city – such issues as platform capacity, parking availability and engine terminal location usually prevent a true 'union terminal' to develop for a large metropolis. Boston has two 'union' terminals for the North and South side commuter lines, while Chicago has four commuter rail terminals.

The solution for a true regional rail system is likely to involve decentralizing the union terminal – both such that the commuter rail market segment could reach a greater number of activity centers directly, and to allow regional rail riders to make an off-on same platform transfer. The design also happens to circumvent the platform capacity issue by requiring all terminating trains to run-through the downtown terminal district, making only 'intermediate' stops and looping back or departing on a different 'spoke' to lay-up in an outlying or suburban engine terminal.

The 'loopback' design is already widespread amongst today's high-density transit operations: Boston's Green Line has a loopback at Government Center and multiple access points along the Central Subway; Chicago's 'el' terminates and run around at the 'Loop'; Philadelphia's Regional Rail system embraces the run-through philosophy. Interurban rail systems that still operate on the push-pull philosophy out of a number of independent downtown terminals are relics of the 20th century and will find themselves unable to service the increasingly decentralized suburb-to-suburb and suburb-to-city-neighborhood transportation demand effectively.

7.1.3 Intercity Rail

Intercity rail, the primary subject of this thesis, is essentially justified on an incremental basis. The capital cost required to create a system of access points in the busy downtown and adjacent city neighborhood activity centers that have gentrified could never be justified by the millions of passengers per year that travel in even the busiest intercity corridors in the United States. Cost of these terminal infrastructure run in the billions, thus have annuities in the region of hundreds of millions of dollars. The convenience of access is simply not worth the extra potentially $100 each passenger would need to pay to have dedicated facilities constructed to enhance intercity access.

However, given how important distributed access is to the commuter rail system, the intercity rail operator simply need to reach agreement with the commuter rail authority to stop trains at key access locations in the city to benefit from the terminal district infrastructure. In drafting such an agreement, it is important to realize that commuter and intercity operators have different priorities and contribute towards the cost of such infrastructure in different ways. The commuter operator is likely to be most concerned about the capacity at the rush-hour, while the intercity operator is probably far more concerned about getting departure slots that are on-the-hour or some other such easy-to-remember number. It is likely that an agreement that calls for shared ownership based on revenues, with provisions that enable the commuter rail operator to take charge of dispatching in the rush hour, while the intercity operator handles the off-peak, would be a sustainable situation in the long run. Any attempt for one party to insist on total control or absolute priority for all trains is likely to fail miserably.

One issue that is likely to occur in intercity rail is the need for exurban Park-n'-Ride lots. The current lots tend to be commuter rail oriented, and is structured to cater towards day-return demands with rates to match. Park-n'-Ride is key to intercity rail's outreach to suburban areas, and the status quo must change when intercity services are extended to more of these facilities.

Another side benefit for having run-through capability in the downtown is that it enables an intercity passenger departing due south to board the train in a northside suburb, ride through the city while the train picks up originating passengers, and depart due south without either having to transfer or drive through the congested city to reach a downtown terminal. The run-through capability is far less critical to the commuter rail market segment than to the regional and intercity rail market segments. However, it is likely that any design that enables access from most of the activity centers in the downtown and inner-city neighborhoods could easily accommodate run-through capability.

7.1.4 Overnight Rail

Once the commuter and intercity infrastructure is in place, accommodating the handful of overnight trains per day is simply a matter of service planning. The driver in service planning in overnight trains is that it must depart within a narrow time window, either between about 9pm to 11pm when passengers are likely to want to begin getting ready for bed, or between about 6pm and 8pm when passengers could board a dinner-and-overnight train. This is less of a problem than at the destination, where the overnight train will by definition arrive in the middle of the morning rush, and will compete for terminal district access slots with commuter rail arrivals.

One way to handle this potential conflict might be to schedule the overnight train much more robustly (say by using the 95th percentile running time) than other trains, or to schedule their arrival towards the end of the morning rush when their impact would be minimal if they were delayed en-route. Needless to say, the passenger from another metropolitan area would need to be delivered to a variety of destinations, and not just the central city – thus the metropolitan access idea is crucial to all market segments.

When dealing with long-distance trains, there is a temptation to cut down the number of stops – especially stops within the same urban area, to minimize running times. This is likely to be a failing strategy. The key driver in the ridership of overnight trains is not the overnight running time, since most passengers would be sleeping and within reason, don't really care how long it takes. The more important factors are departure times, and how many passengers the train can pickup within the narrow time window in which overnight trains could begin. The train may have to stop at a number of suburban or downtown stops to pickup just a handful of passengers, but to require these passengers to travel to a central collection before proceeding is likely to add about an hour or so to their total journey time. Significantly, this hour is far more important than an extra hour overnight – since it is a precious evening hour that the passengers could spend enjoying the city. The likely impact of cutting those extra stops is that those handful of passengers would likely make other travel arrangements. This example demonstrates the importance of customer-centric utility analyses, based on different values-of-time for different parts of the trip and different time-of-day.

7.2 Summary of Important Results

This thesis is a very general and very broad attempt at reframing the high-speed rail debate in the United States. It demonstrates that models are just tools with which the transportation professionals study transportation systems. Just as "Do-Re-Mi and so on are only the tools we use to build music... you can sing a million different tunes by mixing them up", building a transportation system will involve an element of design, and models can serve to inform the impact of different designs on different market segments but should not be the driving force behind choosing certain designs.

Designs are done by artists who take many qualitative factors into account and bring about a vision, while modelling is done by engineers who quantitatively evaluate the impact of the design decisions on passenger utility, project cost, and other indicators that matter. Sometimes, it would be necessary to override the search for an optimum design if other factors such as equity and marketability results in the need to make the system suboptimal in order to serve a broader segment of the population, or achieve other design objectives. By the same token, the fact that a system does not turn a profit is not necessarily an excuse to seek the lowest cost design; in some circumstances a higher cost design will reduce losses by increasing revenues more than the costs.

7.2.1 Modelling Methodologies are Flawed

In Chapter 2, a variety of current approaches to modelling intercity transportation demands are reviewed. The review revealed that even though current methodologies do a good job of predicting incremental changes in passenger demands in response to operating plan changes, addition of new links, and other such minor adjustments, the models do a terrible job of explaining why such changes occur. Consequently, the current models are poor tools with which to inform designers of systems how the system could be altered to both make better use of existing infrastructure and to enhance the customers' intercity travel experience.

Trying to design an intercity transportation system with models of today is like trying to reproduce an elephant by measuring its skin perturbations with a sliderule. The sliderule is unsuitable for the job on two counts: (a) it examines the object with such a coarse resolution (i.e. a straight edge) that it misses subtleties of the elephant, such as the texture of its skin; (b) it examines the object on such a small scale that it fails to realize the right angle at the end of an elephant's belly is part of its leg and isn't because the elephant has been subject to folding action like a piece of paper. Current methodologies, in general, (a) examines the value-of-time too coarsely to appreciate that passengers prefer in-vehicle time spent reading the papers and drinking the coffee to in-vehicle time spent crowded out in a subway car; (b) fails to acknowledge that changes to other parts of the system (such as the addition of a highway link) could affect the part of system under study (such as air travel demand). Although the models are extremely sensitive to aircraft gauge, aircraft schedules, and perhaps even pricing strategies, they are no good for creating strategic visions and answering questions such as: should we widen the interstate highway or connect the high-speed rail to the airport and the suburbs? To answer these questions, a total logistics approach that calculates the total passenger utility by adding all constituent components, similar to the state of practice in freight carrier choice decision support, is required.

7.2.2 Making Journey Time Disappear

It is possible to make journey time "disappear" during long intercity trips. Through of a number of theoretical constructs and review of explicit references to differences in consumer values-of-time while engaged different activities, Chapter 3 demonstrates that it is possible, at least theoretically, to make onerous journey time "disappear" during long intercity trips. The journey time disappears in the sense it does not contribute to the disutility to making the trip. Under certain circumstances, such as when the travellers are sleeping, eating, watching television, or taking part in other household activities that would normally be assigned a zero value-of-time at one's own home, the traveller is indifferent to a longer trip, as long as one would continue to choose to engage that activity and isn't constrained by the limited in-vehicle amenities available. In essence, when boredom sets in, that is when disutility shoots up through the roof. These differences in values-of-time explains the different values-of-time found in studies conducted on otherwise similar people.

Regular users of long distance services have already learned how to make journey time disappear: they chat, bring a deck of cards, or bring reading materials and refreshments. This result borders on stating the obvious, but it is under-appreciated by those who seek to reduce line-haul journey time at the expense of access time. The objective of the high-speed rail advocate should be both to reduce travel time and to make time disappear – and not simply to increase operating speeds.

7.2.3 State Rail Plans are Flawed

In Chapter 4, review of current high-speed rail planning exercise revealed that the planners are examining a gamut of variables that are too narrow. State Rail Plans, or even Federal high-speed rail designated corridors, sometimes define parameters that are too narrow for planners to design a good system. The very act of designating a corridor to be studied could encourage an engineering-approach where the planners come to believe the objective is to find the path of least cost to connect two ill-defined arbitrary end points, such as Boston & Montreal. Firstly, both Boston and Montreal are large metropolitan areas – where is the high-speed rail heading to and from? Secondly, is Boston & Montreal the most logical corridor in the area – what about Boston & Maine, or simply a connection to Manchester Airport? Thirdly, are there other enhancements in the area that would benefit the locality more than a high-speed rail – what about the North South Rail Link, or simply a number of intermodal passenger terminals coupled with medium-speed rail service? These are issues needing to be addressed in system design, and do not seem to have been addressed in the high-speed rail plans reviewed.

7.2.4 Metropolitan Access is Vital

In Chapter 5, this activity-dependent disutility framework (or total passenger logistics-utility) was extended to model a series of rail terminal locations, layouts, and service designs in large metropolitan areas. There are two main results. One main realization was that not only does the activity drive the value-of-time, the distribution of time available for activities are also important – simply by changing the mix of times can reduce "logistics costs", or disutility, for many passengers. Another finding is that improved access can be an effective way of reducing total trip disutility – making the trip faster by cutting transfer time, and also more productive and more comfortable.

The first result largely reflects the idea that people do not like to be interrupted from their task, whether it is work- or leisure-related. Contrary to conventional wisdom in transit environments, a one-seat-ride could be an important factor in intercity mode choice. Each segment in an intermodal intercity itinerary or a longer commute tend to be of a length where productive work is possible, the passenger would have a stronger preference for uninterrupted time than a transit-rider.

Applying a version of the logistics-utility model, the City of London case study shows that investing in better access to the downtown core could save much more time for passengers than investing in increasing the speeds of commuter lines radiating from London. In the scenario studied, average journey time savings (for all origins on the former Network Southeast system) could be as much as ten minutes compared to a transfer to the London Underground Circle Line. The results are highly intuitive: journey time reductions on the line-haul segments tend to help the outer suburbs, where the demands are low, whilst enhancing access to the core help the inner suburbs, the outer suburbs, as well as crosstown passengers. Again, those who focus on one single origin-destination pair or one single corridor often ignore these important results. Intercity rail is a network, even if the corridor may seem as simple as a beeline between one city and the next – there is always a cluster of metropolitan neighborhoods that need to be connected with another cluster, and any evaluation must consider all origin-destination pairs.

Applying the total logistics-utility model to services between two smaller cities resulted in a design that calls for an intercity line that winds around the city to collect passengers from multiple neighborhoods in a linear fashion. This is the MetroFlyer concept, introduced in Chapter 5. In smaller cities, where there are insufficient demand density to justify a totally-connected network from every neighborhood to every other, it is possible to construct such a line to achieve better access before heading out towards another city or the suburbs. This type of infrastructure, which can sometimes be constructed out of abandoned and pre-existing rail lines, can benefit the city in a number of ways: (1) more city neighborhoods will receive regional and intercity rail access, (2) new corridors are created where a subway-like service may relieve capacity problems on the existing transit system, (3) intermodal and intramodal connections could be made at multiple nodes or multiple 'union stations', which will alleviate the congestion and parking problem typically associated with union stations.

The total logistics concept for passengers gives rise to an important methodology with which high-speed rail schemes should be evaluated. Firstly, the issue of access to high speed rail services has to be taken seriously, if high-speed rail advocates wishes to be considered as a viable voice in promoting mass intercity transportation. Secondly, to demonstrate that a high speed is really necessary, the speed enhancement must be explicitly traded off against other possible enhancements in a utility framework.

7.3 High Speed Rail Planning Synthesis

7.3.1 Regarding the Value of Time

The logical consequence of acknowledging the activity-dependent disutility of in-vehicle time is that the standard model of disutility equals the product of time and value-of-time will no longer apply. Instead, the disutility should be modelled as the integral of values-of-time over the entire trip, from the moment the travellers leave their point of origin until they reach their destination. Access time, terminal time, buffer time, in-vehicle time are all included, and the values-of-time in each category would still require careful evaluation – for instance, terminal time spent browsing through the concessions has a different value from terminal time spent standing in line waiting for a security check-in. These are factors that designers of intercity transportation systems must pay careful attention to. Once the infrastructure is built, the users may have already been trapped into a suboptimal path from the point of view of total trip utility.

Amenities may be expensive to provide from the carrier's perspective, but they are part of the carrier's competitive arsenal. Highway amenities are provided by independent businesses on a commercial basis, and airport amenities are provided by concessions. Rail carriers can do well by explicitly recognizing the link between amenities and value-of-time en-route. While it is not necessary to provide free amenities, they should be provided at a cost comparable to similar amenities at airports or highway rest-stops, to encourage the travellers to make their own journey time disappear.

7.3.2 Regarding High Speed Rail Planning in General

- Current models in intercity transportation often consider only one aspect of the broader intercity transportation problem. Airline demand models may deal with such issues as carrier choice and schedule choice, but do not address issues such as local access and terminal amenities -- nor were they designed to. Strategic planners in intercity transportation should not confuse the need for transportation systems design with short-term operations planning models that were essentially designed to evaluate incremental costs and benefits of improving the operating plan.

- Most of the important concepts in designing an intercity transportation system already exist in the literature. Total logistics costs as appled to freight transportation, mode-choice methodologies as applied to urban transportation, and even explicit considerations of customers' value of time, are all concepts that can assist the evaluation of different designs of intercity transportation systems. However, the existence of modelling methodologies does not alleviate the need for design. Design is a separate craft in which the professionals gather inputs from multiple stakeholder groups and combine them in such a way as to make a functional transportation system.

- By the same token, designing a transportation system is not a simple process, and may not relate to available transportation technologies as strongly as often suggested. The availability of a new technology will influence the design process, since new technologies may cater to certain users and stakeholders' interests better than older technologies. However, designing a good system is more than simply taking a new technology, building and calibrating a model to show an instance where the new technology would work better than the old. Only when much thought had been devoted to how best to utilize the new technology and how it affects role of the older technologies, should the whole system be evaluated using a model to show that the deployment will benefit many different stakeholders. In particular, it is important to avoid situations where new technologies may do well at the expense of an older part of the system -- good design would put existing infrastructure to good use while allowing the new technology to serve a useful purpose.

- In essence, a demand model that demonstrates that there is sufficient demand to justify financially the operations of a new corridor, a new technology, or some piece of new infrastructure, is not necessary and sufficient to justify its construction. A vision, a design proposal, coupled with a systems evaluation of what the new infrastructure will do to the users and the non-users, is much more important than a narrowly focused study that simply claims 'new stuff is needed here, and it will make money'. This may seem obvious, but many state rail plans and other strategic plans for intercity rail or airports appear to overlook the need to examine the transportation problem on a systems level.

- Much expertise has been developed in urban transportation systems design. The Boston Metropolitan Transportation Plan (1972) captured the essence of the kind of design considerations that were needed. The plan integrated for the first time a proposal for new highways, new subways, and other transportation facilities in the area. When part of the plan was implemented, new housing was built along the new subway alignment to create a 'livable neighbourhood'. This type of attention to detail, and systems approach to planning, is needed in intercity transportation planning.

- The systems approach to transportation planning often calls for intermodal transportation connexions. Intermodal connexion is only one of many ways to achieve a systems vision. Integration of airports, commuter rail, and intercity rail can be important, but should not rule out the possibility of constructing infrastructure in such a way that a service can transcend many different modal roles. For instance, in the terminal areas, an intercity arrival could turn into a commuter train as it approaches an urban area; local trains should travel by different routes from intercity trains where possible, to provide maximum connectivity; airlines should focus on what they do best -- service isolated cities, and provide an ultra high-speed service for those who are willing to pay for it.

These conclusions are somewhat broad. An immediate possibility is to re-evaluate some of the proposals currently in progress in light of the metropolitan access findings, taking into account total trip time for all likely customers, and try to evaluate the value of time. It is likely that some changes will be needed to bring the proposals to a stage where it better benefits the regional transportation system as a whole, instead of being an isolated corridor.

7.4 Recommendations and Future Work

As indicated throughout this thesis, much remains to be done to develop a framework for planning intercity rail transportation systems. The current institutions involved in intercity transportation have tended to be modal-specific, sometimes resulting in mal-coordinated systems. In urban areas, because of the massive amount of public subsidy that has been poured into public transportation, methodologies have been developed for coordinating the services provided by different modes, and catering for the different needs of the urban market though joint-planning and regulation. Although deregulation of intercity air carriers was hailed as a economic success, it has resulted in reinforcement of the modal mentality – only in the case of one carrier, has the issue of joint highway-air service been considered seriously. There is now a great need for each metropolitan area to examine their intercity service facilities: are the passengers getting from their origins to the destination city by the most efficient route through the urban area? The chances are that most passengers are taking a geographic detour because they have to travel to the airport, and may even be suffering time penalties in the case of shorter-distance origin-destination markets.

In terms of analyzing the economics of schemes such as MetroFlyer, and other urban distribution systems for intercity carriers, location specific studies are needed to examine the costs and benefits associated with alternatives for a specific location. The problem with some state passenger rail plans lies in not examining all possible alternatives, either because the MetroFlyer alternative is not initially obvious, or other factors are preventing it from being considered. If the systems analysis presented in this thesis is even halfway correct, in many metropolitan areas it will be demonstrated that improving access to intercity and commuter rail facilities will benefit more passengers (and save more passenger-minutes) than simply improving speed for a specific origin and destination.

Plate 7.2: Considering the MetroFlyer alternative would not only help cities like New York, but also regional cities that could justify a trunk distributor through the metro area.

7.4.1 The Speed Assumption

If there were a moral to this thesis, it would be to avoid the 'speed assumption' in future work in intercity transportation planning. To demonstrate that a high speed is really necessary, the speed enhancement must be explicitly traded off against other possible enhancements in a utility framework. Speed, like any other amenity, requires justification with sound project evaluation. Frequently, speed could be justified at the margin – like in the Tokaido Shinkansen; in many cases however, a lot could be accomplished with sound service planning and by exploiting synergies with parts of the public transportation system with much higher ridership, such as the commuter rail.

Currently in the United States, higher speed rail is necessary in many cities for rail to stay competitive, but highest speed rail is probably neither cost-effective nor necessary. Instead, each scheme for increasing line-haul speed should be judged, using the total logistics-utility framework, against a series of alternatives to improve access to locations of large demand density as well as options that help to make time disappear. In the same vein, more accessible rail offering shorter access time from all points in the city is more and more important in today's sprawling metropolitan areas. However, the most accessible rail (such a streetcar) that calls at every street corner, have little role in interurban transportation. Demands for speed, accessibility, amenities, and other upgrades that improve the customer utility must be balanced against each other. The results from the customer utility studies should be used to inform intercity transportation system design, to create a system that works in harmony to move people.

Appendices

Metropolitan Access,
Intercity Rail and Technology